

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2000, 18:38:42 ; Search time 6423.57 Seconds
(without alignments)
2513.359 Million cell updates/sec

Title: US-09-090-672B-1
Perfect score: 4276
Sequence: 1 TTCTACCGTTTTTCCCTGC.....ATCAGAAAAAAAAAAAAA 4276

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 50.0

Searched: 4538634 seqs, 1887031982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	386	9.0	506	45	AI392674 tg47c02.x

2	348	8.1	398	24	H71225	H71225 ys12e09.r1
3	329	7.7	438	24	H71226	H71226 ys12e09.s1
C 4	322	7.5	529	38	AA772278	AA772278 ai42d03.s
C 5	320	7.5	329	20	Z19960	Z19960 HSAABDCI.B
C 6	302	7.1	302	33	AA381126	AA381126 EST94221
C 7	244	5.7	364	36	AA634469	AA634469 zu76a06.s
C 8	235	5.5	395	60	AI802170	AI802170 tx25b02.s
C 9	193	4.5	274	38	AA782211	AA782211 ai47h12.s
C 10	191	4.5	221	39	AA903370	AA903370 OK58C02.S
C 11	188	4.4	201	33	AA417896	AA417896 zv05e07.s
C 12	187	4.4	385	21	T93967	T93967 y056d10.s1
C 13	186	4.3	365	42	AI093858	AI093858 qa30d05.s
C 14	175	4.1	433	63	AI962924	AI962924 wt24h06.x
C 15	156	3.6	156	33	AA381125	AA381125 EST94220
C 16	140	3.3	410	23	H57530	H57530 yr05b10.r1
C 17	135	3.2	559	88	AQ883641	AQ883641 HS_5478.B
C 18	130	3.0	304	21	T85805	T85805 yd56d10.r1
C 19	93	2.2	568	51	AI762436	AI762436 wq57a05.x
C 20	53	1.2	639	81	B55278	B55278 CIT-HSP-386
C 21	47	1.1	764	87	AQ744981	AQ744981 HS_5501.A
C 22	46	1.1	261	90	AQ061635	AQ061635 CIT-HSP-2
C 23	46	1.1	269	32	AA350648	AA350648 EST58017
C 24	46	1.1	343	40	AA955031	AA955031 oq66h02.s
C 25	46	1.1	434	105	AQ608513	AQ608513 HS_2124.B
C 26	45	1.1	426	22	R50700	R50700 Y360g12.r1
C 27	44	1.0	534	102	AQ388631	AQ388631 RPI111-14
C 28	44	1.0	602	81	B58317	B58317 CIT-HSP-201
C 29	43	1.0	180	36	AA654395	AA654395 nt03e09.s
C 30	43	1.0	383	62	AI926033	AI926033 w041D09.x
C 31	43	1.0	391	28	AA121819	AA121819 zn35g09.r
C 32	43	1.0	406	184	AQ315717	AQ315717 HS_5235.A
C 33	43	1.0	470	49	AL047553	AL047553 DKEP386E
C 34	42	1.0	527	102	AQ394316	AQ394316 CITBI-E1-
C 35	41	1.0	389	99	AQ205998	AQ205998 HS_3237-B
C 36	41	1.0	400	99	AQ206027	AQ206027 HS_3237-B
C 37	41	1.0	513	82	AQ763211	AQ763211 HS_3222-B
C 38	41	1.0	693	82	AQ742061	AQ742061 HS_5570-B
C 39	39	0.9	104	100	AQ268072	AQ268072 RPI111-73
C 40	39	0.9	274	27	AA054055	AA054055 zf48e07.f
C 41	39	0.9	301	91	AQ100978	AQ100978 HS_3081.A
C 42	38	0.9	115	28	AQ084176	AQ084176 zn17d10.s
C 43	38	0.9	132	24	H84675	H84675 ys66d04.r1
C 44	38	0.9	151	40	AA908331	AA908331 og75d06.s
C 45	38	0.9	157	33	AA397419	AA397419 nc64f11.s

ALIGNMENTS

RESULT 1
 AI392674/c
 LOCUS
 DEFINITION
 t947c02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:211906 3', mRNA sequence.
 AI392674
 AI392674.1 GI:4222221
 EST.
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 506)
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Jan 19, 1998 this sequence version replaced gi:2151612.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Gibco

High quality sequence stop: 444.
 Location/Qualifiers
 1. 506
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:211906"
 /lab_host="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7f3D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI_CGAP_GCS1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 582632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo.
 BASE COUNT 185 a 75 c 82 g 164 t
 ORIGIN
 Query Match 9.0%; Score 386; DB 45; Length 506;
 Best Local Similarity 99.8%; Pred. No. 1.2e-136;
 Matches 506; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 3762 ACAATATGAATCTCTTTTCGTATGCCATCGGTGCGATGGAAGTTTATCTCTGTT 3821
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 506 ACAATATGAATCTCTTTTCGTATGCCATCGGTGCGATGGAAGTTTATCTCTGTT 447
 QY 3822 TTCTGGAACCAAGAGATGCCAACTCTCTGCAACATTTCTTAGAGGAGAGAGAAA 3881
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 446 TTGCTGGAACCAAGAGATGCCAACTCTCTGCAACATTTCTTAGAGGAGAGAGAAA 387
 QY 3882 TATTAAGAGAAATGAACAATAGAGTATTTGGGTTTTTAATTAATTAATTTGTTAATA 3941
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 386 TATTAAGAGAAATGAACAATAGAGTATTTGGGTTTTTAATTAATTAATTTGTTAATA 327
 QY 3942 ATATTAACATATAAGATACATTTTATTAATAACATGCAACAATAACACATATCGGCTCA 4001
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 326 ATATTAACATATAAGATACATTTTATTAATAACATGCAACAATAACACATATCGGCTCA 267
 QY 4002 TCTGACAGTTTTTCCCGCCAGGAGTGCTTTTGCCTTTTCCCTTTCTTTTCTTTTCTTTT 4061
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 266 TCTGACAGTTTTTCCCGCCAGGAGTGCTTTTGCCTTTTCTTTTCTTTTCTTTTCTTT 207
 QY 4062 CATCTTTTCTCTCTCTCTTTTTCCTATCCCTTTTAAATTTTAAACAGCAATGGAG 4121
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 206 CATCTTTTCTCTCTCTCTTTTTCCTATCCCTTTTAAATTTTAAACAGCAATGGAG 148
 QY 4122 GAAGTTAAACAATTTTAAATGGAAGAGCATGTTAGAGCAACAATAAGCAAGACT 4181
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 147 GAAGTTAAACAATTTTAAATGGAAGAGCATGTTAGAGCAACAATAAGCAAGACT 88
 QY 4182 GAGCAGCATATAATTAATTTTCAGGGTTTTGAGGCTGCAACATAATTTTCATTATCCCTCA 4241
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 87 GAGCAGCATATAATTAATTTTCAGGGTTTTGAGGCTGCAACATAATTTTCATTATCCCTCA 28
 QY 4242 AAAAGTTTACCACCATCATGAGAAAAA 4268
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 27 AAAAGTTTACCACCATCATGAGAAAAA 1
 RESULT 2
 H71225
 LOCUS
 DEFINITION
 ys12e09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
 IMAGE:214600 5' similar to SP.C4081.1 CE00109 OVARIAN PROTEIN ;,
 mRNA sequence.
 H71225
 H71225.1 GI:1043041
 ACCESSION
 VERSION

Qy 3550 CCCTCGCATATTTCAATTCGCTGGAAGTAAGGATCACTGCTCATTTTCAGGCC 3609
 Db 329 CCCTCGCATATTTCAATTCGCTGGAAGTAAGGATCACTGCTCATTTTCAGGCC 270
 Qy 3610 TCAGAATAAGTGCACCTCTTCTGTTTCATCTGACCCCTTCCTCAACCTCTTCACGCTGGCA 3669
 Db 269 TCAGAATAAGTGCACCTCTTCTGTTTCATCTGACCCCTTCCTCAACCTCTTCACGCTGGCA 210
 Qy 3670 TGTCCTTTTGTACAGCTCTGTAACTTAAGTATAGTATATGAAAGAAATGACCTATAATA 3729
 Db 209 TGTCCTTTTGTACAGCTCTGTAACTTAAGTATAGTATATGAAAGAAATGACCTATAATA 150
 Qy 3730 TAGGTGTTTGTAGATCTTCTGCTCACTGCAACAAATGAACTCCCTTTTCGTATGTC 3789
 Db 149 TAGGTGTTTGTAGATCTTCTGCTCACTGCAACAAATGAACTCCCTTTTCGTATGTC 90
 Qy 3790 ATCGGGTGCATGGAAGTTTATTCCTGTTGTTGCTGGAACCAAGAGGATCCAAACTT 3849
 Db 89 ATCGGGTGCATGGAAGTTTATTCCTGTTGTTGCTGGAACCAAGAGGATCCAAACTT 30
 Qy 3850 CCTGCAACATTTCTTAGAGGAGAGAGAG 3878
 Db 29 CCTGCAACATTTCTTAGAGGAGAGAG 1

RESULT 4
 AA772278/c
 LOCUS
 DEFINITION
 accession
 version
 keywords
 source
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 529)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On May 18, 1995 this sequence version replaced gi:811190.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1133 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 475.
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="1359653"
 /clone_lib="Soares_parathyroid_tumor_NbHPA"
 /tissue_type="parathyroid tumor"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /note="organ: parathyroid gland; Vector: pT73D
 (Pharmacia) with a modified polylinker; Site.1: Not I;
 Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 15'-
 TGTACCATCTGAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTTTT

FEATURES
 source

BASE COUNT 196 a 86 g 167 t
 ORIGIN

Query Match 7.5%; Score 322; DB 38; Length 529;
 Best Local Similarity 100.0%; Pred. No. 2.4e-112; Indels 0; Gaps 0;
 Matches 322; Conservative 0; Mismatches 0;

Qy 3740 GTAGATTCCTGTGCTGCAACAAATATGAACCTCTTTTCGTATGCGATCGGGTTC 3799
 Db 529 GTAGATTCCTGTGCTGCAACAAATATGAACCTCTTTTCGTATGCGATCGGGTTC 470
 Qy 3800 ATGGAAGTTTATTCCTGTTGTTGCTGGAACCAAGAGGATCCAAACTTCCTGCAACAT 3859
 Db 469 ATGGAAGTTTATTCCTGTTGTTGCTGGAACCAAGAGGATCCAAACTTCCTGCAACAT 410
 Qy 3860 TTCTCTTAGAGGAGAGAGAGAAATATTAAGAGAGAAATGAACAAATAGAGTATTTGGGTT 3919
 Db 409 TTCTCTTAGAGGAGAGAGAGAAATATTAAGAGAGAAATGAACAAATAGAGTATTTGGGTT 350
 Qy 3920 TTTAATAATATTTGTTTAATAATATTAACATATGAAGTACTTTTAAATAAACCATG 3979
 Db 349 TTTAATAATATTTGTTTAATAATATTAACATATGAAGTACTTTTAAATAAACCATG 290
 Qy 3980 CAACAATAACACTATCGGTCTATCTGACAGTTTTTCCGCCAGGAGAGTCTTTGCTTT 4039
 Db 289 CAACAATAACACTATCGGTCTATCTGACAGTTTTTCCGCCAGGAGAGTCTTTGCTTT 230
 Qy 4040 TCCTTCTTTTCTTTTCTTTTCTTTT 4061
 Db 229 TCCTTCTTTTCTTTTCTTTTCTTTT 208

RESULT 5
 Z19960/c
 LOCUS
 DEFINITION
 accession
 version
 keywords
 source
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 329)
 The UK-HGMP cDNA program
 Unpublished (1993)
 Contact: MRC Human Genome Mapping Project Resource Centre
 Clinical Research Centre
 Watford Road, Harrow, Middlesex HA1 3UJ, U.K.
 Email: biohelp@hgmp.mrc.ac.uk
 Single read.
 Location/Qualifiers
 1..329
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="B, Human Liver tissue"
 /note="vector: gtl1; clone_library=B, Human Liver tissue;
 cloning vector is gtl1."
 110 a 56 c 58 g 105 t
 BASE COUNT
 ORIGIN

T-3'], double-stranded cDNA was size selected, ligated to
 Eco RI adapters (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
 adenomas was kindly provided by Dr. Stephen Marx, National
 Institute of Diabetes and Digestive and Kidney Diseases,
 NIH."

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 299.

FEATURES

source
1. .364
/organism="Homo sapiens"
/db_xref="GDB:5931274"
/db_xref="taxon:9606"
/clone="IMAGE:743890"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 142 a 47 c 50 g 115 t
ORIGIN

Query Match 5.7%; Score 244; DB 36; Length 364;
Best Local Similarity 99.7%; Pred. No. 1.2e-82;
Matches 364; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3898 AACCAATAGAGTATTTGGGTTTTTAATTAATTTGTTAATATACATATAAGAA 3957
Db 364 AACCAATAGAGTATTTGGGTTTTTAATTAATTTGTTAATATACATATAAGAA 305
QY 3958 TACTTTTATTAATAAACATGCAACATAACACTATCGGTCTATCTGACATTTTCC 4017
Db 304 TACTTTTATTAATAAACATGCAACATAACACTATCGGTCTATCTGACATTTTCC 245
QY 4018 CCAGGAGAGTGTTCCTTCCTTCCTTTTCCTTTTCCTTTTCCTTTTCCTTC 4077
Db 244 CCAGGAGAGTGTTCCTTCCTTCCTTTTCCTTTTCCTTTTCCTTTTCCTTC 186
QY 4078 TCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAAGTTACAAATTTT 4137
Db 185 TCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAAGTTACAAATTTT 126
QY 4138 AATGGAAGAGCATGTTAGACACAAATGCATAGCAGACTGAGCAGCATTTATTT 4197
Db 125 AATGGAAGAGCATGTTAGACACAAATGCATAGCAGACTGAGCAGCATTTATTT 66
QY 4198 AATTTTCAGGGTTTGGGCTGAACATAAATTCATTATCCCTCAAAAAGTTACCA 4257
Db 65 AATTTTCAGGGTTTGGGCTGAACATAAATTCATTATCCCTCAAAAAGTTACCA 6
QY 4258 TCAGA 4262
Db 5 TCAGA 1

RESULT 8
AI802170
LOCUS AI802170 395 bp mRNA EST 06-JUL-1999
DEFINITION tx25b02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2270571 3', mRNA sequence.
ACCESSION AI802170
VERSION AI802170.1 GI:5367642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 395)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1131028.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
CDNA Library Arrayed by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .395

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2270571"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu24 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 130 a 58 c 61 g 146 t
ORIGIN

Query Match 5.5%; Score 235; DB 60; Length 395;
Best Local Similarity 100.0%; Pred. No. 3e-79;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3827 GGAACCAAGAGATCCAACTTCCTGCAACATTTCTTAGAGGAGAGAGAAATATTA 3886
Db 1 GGAACCAAGAGATCCAACTTCCTGCAACATTTCTTAGAGGAGAGAGAAATATTA 60
QY 3887 AAAGAGAAATGAACAATAGATTTTGGGTTTAAATTAATTAATTAATTAATTA 3946
Db 61 AAAGAGAAATGAACAATAGATTTTGGGTTTAAATTAATTAATTAATTAATTA 120
QY 3947 ACATATAAGAAATCTTTTAAATAACCATGCAACATAACACTATCGTCTATCTGA 4006
Db 121 ACATATAAGAAATCTTTTAAATAACCATGCAACATAACACTATCGTCTATCTGA 180
QY 4007 CAGTTTTTCCCCCAGGAGAGTGTTCCTTCCTTCCTTTCTTTCTTTCTTTT 4061
Db 181 CAGTTTTTCCCCCAGGAGAGTGTTCCTTCCTTTCTTTCTTTCTTTCTTTT 235

RESULT 9
AA782211/c
LOCUS AA782211 274 bp mRNA EST 31-DEC-1998
DEFINITION a147h12.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone 1360199 3', mRNA sequence.
ACCESSION AA782211
VERSION AA782211.1 GI:2841542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE
AUTHORS      1 (bases 1 to 274)
TITLE        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
SOURCE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
COMMENT      On Sep 29, 1997 this sequence version replaced gi:1520658.
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bonaldo, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 1172 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 253.

FEATURES
Source
1..274
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="1360199"
   /clone_lib="Soares_parathyroid_tumor_NbHPA"
   /tissue_type="parathyroid tumor"
   /dev_stage="adult"
   /lab_host="DH10B"
   /note="Organ: parathyroid gland; Vector: pT7T3D
   (Pharmacia) with a modified polylinker; Site_1: Not I;
   Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
   oligo(dT) primer
15'-
TGTACCAATCTGAAGTGGGAGCGCCGACCAATTTTTTTTTTTTTTTTTT
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
BASE COUNT      104 a      39 c      54 g      77 t
ORIGIN

Query Match
Best Local Similarity 4.5%; Score 193; DB 38; Length 274;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4074 TCTCTCTCTTTTTCATCCCTTTTAAATTTTTTAAACAGCAATGGAGGAAGTTAACAAAT 4133
      |||
Db 193 TCTCTCTCTTTTTCATCCCTTTTAAATTTTTTAAACAGCAATGGAGGAAGTTAACAAAT 134
      |||

QY 4134 TTTTAATGGAAGAGCATGTTAGAGCAACAAATGCAATGAAGCAAGCTGAGCAGCATTAAT 4193
      |||
Db 133 TTTTAATGGAAGAGCATGTTAGAGCAACAAATGCAATGAAGCAAGCTGAGCAGCATTAAT 74
      |||

QY 4194 AATTAATTTTCAGGGTTTGAGGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 4253
      |||
Db 73 AATTAATTTTCAGGGTTTGAGGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 14
      |||

QY 4254 CACATCAGAAAA 4266
      |||
Db 13 CACATCAGAAAA 1
      |||

RESULT 10
AA903370/c 221 bp mRNA EST 24-AUG-1998
LOCUS OK58c02.s1 NCI_CGAP_Le12 Homo sapiens cDNA clone IMAGE:1518146 3',
DEFINITION
```

```
mrna sequence.
AA903370 GI:3038493
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 221)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1398114.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
unknown library type
Insert Length: 1536 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 178.

FEATURES
Source
1..221
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:1518146"
   /clone_lib="NCI_CGAP_Le12"
   /tissue_type="leiomyosarcoma"
   /lab_host="DH10B"
   /note="Organ: soft tissue; Vector: pT7T3D-Pac (Pharmacia)
   with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
   1st strand cDNA was primed with a Not I - oligo(dT) primer
   15'-AACTGGAGAGATTCGCGCCGCAATGTTTTTTTTTTTTTTTTT-3'],
   double-stranded cDNA was ligated to Eco RI adaptors
   (Pharmacia), digested with Not I and cloned into the Not I
   and Eco RI sites of the modified pT7T3 vector. Library
   went through one round of normalization. Library
   constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      87 a      29 c      38 g      67 t
ORIGIN

Query Match
Best Local Similarity 4.5%; Score 191; DB 39; Length 221;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4074 TCTCTCTCTTTTTCATCCCTTTTAAATTTTTTAAACAGCAATGGAGGAAGTTAACAAAT 4133
      |||
Db 191 TCTCTCTCTTTTTCATCCCTTTTAAATTTTTTAAACAGCAATGGAGGAAGTTAACAAAT 132
      |||

QY 4134 TTTTAATGGAAGAGCATGTTAGAGCAACAAATGCAATGAAGCAAGCTGAGCAGCATTAAT 4193
      |||
Db 131 TTTTAATGGAAGAGCATGTTAGAGCAACAAATGCAATGAAGCAAGCTGAGCAGCATTAAT 72
      |||

QY 4194 AATTAATTTTCAGGGTTTGAGGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 4253
      |||
Db 71 AATTAATTTTCAGGGTTTGAGGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 12
      |||

QY 4254 CACATCAGAAAA 4264
      |||
Db 11 CACATCAGAAAA 1
      |||

RESULT 11
AA417896/c 201 bp mRNA EST 02-MAR-1998
LOCUS zv05e07.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:752772
DEFINITION 3', mRNA sequence.
ACCESSION AA417896
VERSION AA417896.1 GI:2079715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```


REFERENCE
AUTHORS
Eutheria: Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 156)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
White, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
Bult, C.J., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, I.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrle, A., Fischer, C., Hastings, G.A.,
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
12140200
On Sep 12, 1996 this sequence version replaced gi:1398060.
Other ESTs: B5794221
Contact: Kerlavage, AR
Bioinformatics for Genomic Research
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018599056
Fax: 3018599423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13-21.
Location/Qualifiers
1..156
/organism="Homo sapiens"
/db_xref="ATCC (inhost):185552"
/db_xref="taxon:9606"
/clone_lib="Activated T-cells I"
/cell_type="T-lymphocyte"
/dev_stage="adult"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 43 a 28 c 25 g 60 t
ORIGIN
Query Match 3.6%; Score 156; DB 33; Length 156;
Best Local Similarity 100.0%; Pred. No. 4.2e-49;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4106 TTTACAGCAATGAGGAGTACAAATTTTATGGAAGAGCATGTTAGGCAACAA 4165
Db 156 TTTACAGCAATGAGGAGTACAAATTTTATGGAAGAGCATGTTAGGCAACAA 97
Qy 4166 ATGCATACAGAGCTGAGCAGCATATATATTTTCAGGCTTTTTCAGGCTGACATA 4225
Db 96 ATGCATACAGAGCTGAGCAGCATATATATTTTCAGGCTTTTTCAGGCTGACATA 37
Qy 4226 ATTTCATTATCCCTCAAAAAGTTACCCACCATCAG 4261
Db 36 ATTTCATTATCCCTCAAAAAGTTACCCACCATCAG 1
RESULT 16
H57530
LOCUS
DEFINITION
Yr03b10_r1 Soares fetal liver spleen 1NfLS Homo sapiens cDNA clone
IMAGE:204379.5' similar to SP:C40H1.1 CE00109 OVARIAN PROTEIN ;,
mRNA sequence.

ACCESSION H57530.1 GI:1010362
VERSION H57530.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On Jan 24, 1995 this sequence version replaced gi:634282.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 902
High quality sequence stops: 361
Source: IMAGE Consortium; LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 902 Std Error: 0.00
Seq primer: M13Rp1
High quality sequence stop: 361.
Location/Qualifiers
1..410
/organism="Homo sapiens"
/db_xref="GDB:3773510"
/db_xref="taxon:9606"
/clone="IMAGE:204379"
/clone_lib="Soares fetal liver spleen 1NfLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(AT) primer
[5' AACTGGAAGATTAATTAAGATCTTTTATTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 97 a 79 c 116 g 115 t
ORIGIN
Query Match 3.3%; Score 140; DB 23; Length 410;
Best Local Similarity 100.0%; Pred. No. 3.5e-43;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 575 TAGCTGGGATGATACCGACCTGACCTAAATACCCAAAGAGCTGGGAGATTCG 634
Db 170 TAGCTGGGATGATACCGACCTGACCTAAATACCCAAAGAGCTGGGAGATTCG 229
Qy 635 TTCTCTAATCAACAGAGTTACATAGCTGCTATCAGTCCCGCTTTGTTTCAGCTGACAT 694
Db 230 TTCTCTAATCAACAGAGTTACATAGCTGCTATCAGTCCCGCTTTGTTTCAGCTGACAT 289
Qy 695 GGAGATAGATAAAGCGGT 714
Db 290 GGAGATAGATAAAGCGGT 309
RESULT 17
A0883641/C
LOCUS
A0883641 559 bp DNA GSS 09-NOV-1999

AI762436/c 568 bp mRNA EST 24-JUN-1999
 LOCUS wg57a05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2369168.3, mRNA sequence.
 ACCESSION AI762436
 VERSION AI762436.1 GI:5178103
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 568)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3187416.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Gibco
 High quality sequence stop: 483.
 FEATURES
 Location/Qualifiers
 1..568
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2369168"
 /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSF pool 1:
 30384-310919, 323208-325893 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NBHFA pool 1:
 758280-760583, 772104-774407 Soares NBHFA pool 1:
 304776-306311, 320136-322823, 326280-326663 Soares NBHOT
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 230 a 79 c 88 g 170 t 1 others
 ORIGIN
 Query Match 2.2% Score 93; DB 51; Length 568;
 Best Local Similarity 100.0%; Pred No. 2.2e-25;
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 205 AGATGAGATCAGCTAGTTTCCTGCTTGGCCCTCTGATTGGATTGGCTCTATAA 264
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 409 AGATGAGATCAGCTAGTTTCCTGCTTGGCCCTCTGATTGGATTGGCTCTATAA 350
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 265 AGCTGAGACCAATCTATTTCTCTCTCTAAAGG 297
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 349 AGCTGAGACCAATCTATTTCTCTCTCTAAAGG 317
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 RESULT 20
 B55278/c 639 bp DNA GSS 20-JUN-1998
 LOCUS CIT-HSP-386K14.TR CIT-HSP Homo sapiens genomic clone 386K14,
 DEFINITION genomic survey sequence.
 ACCESSION B55278
 VERSION B55278.1 GI:2609612
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 764)
 AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 639)
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: CIT-HSP-386K14.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: M13 Reverse
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 1..639
 /organism="Homo sapiens"
 /db_xref="GDB:5379185"
 /db_xref="taxon:9606"
 /clone="386K14"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII"
 BASE COUNT 197 a 141 c 159 g 142 t
 ORIGIN
 Query Match 1.2% Score 53; DB 81; Length 639;
 Best Local Similarity 100.0%; Pred No. 3.2e-10;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1894 CAACCTCTCCCTCCGGTTCAAGTGATTCCTCTGCTTAGCTCCCGAGTAG 1945
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 CAACCTCTCCCTCCGGTTCAAGTGATTCCTCTGCTTAGCTCCCGAGTAG 309
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 RESULT 21
 AQ744981/c 764 bp DNA GSS 16-JUL-1999
 LOCUS HS_5501.A2.C03.SP6.RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-1077 Col-6 Row-E, genomic survey sequence.
 ACCESSION AQ744981
 VERSION AQ744981.1 GI:5522591
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 764)
 AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 1077 row: E column: 6
 Seq primer: Sp6
 Class: BAC ends
 High quality sequence stop: 764.

Location/Qualifiers

1..764
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1077 Col=6 Row=E"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT 233 a 138 c 164 g 228 t 1 Others

ORIGIN

Query Match 1..18; Score 47; DB 87; Length 764;
 Best Local Similarity 100.0%; Pred. No. 5.7e-08; Indels 0; Gaps 0;
 Matches 47; Conservative 0; Mismatches 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1940
 |||||
 Db 422 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 376

RESULT 22

LOCUS

DEFINITION A0061635 261 bp DNA GSS 31-JUL-1998
 CIT-HSP-234805.TF CIT-HSP Homo sapiens genomic clone 234805,
 genomic survey sequence.

ACCESSION A0061635

VERSION A0061635.1 GI:3363547

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 261)
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M. and Venter,J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)
 Unpublished
 Other_GSSs: CIT-HSP-234805.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="234805"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /note="vector: pBelobAC11; Site_1: HindIII; Site_2:
 similar to contains Alu repetitive element;contains element MER22

BASE COUNT 42 a 62 c 66 g 91 t

ORIGIN

Query Match 1..18; Score 46; DB 90; Length 261;
 Best Local Similarity 100.0%; Pred. No. 2.1e-07;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1939
 |||||

Db 107 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 152

RESULT 23

LOCUS

DEFINITION AA350648 269 bp mRNA EST 21-APR-1997
 EST58017 Infant brain Homo sapiens cDNA 3' end similar to EST
 containing Alu repeat, mRNA sequence.

ACCESSION AA350648

VERSION AA350648.1 GI:2002965

KEYWORDS EST.

SOURCE human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 269)
 Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
 3,400 expressed sequence tags identify diversity of transcripts
 from human brain
 Nature Genet. 4, 256-267 (1993)
 9336420
 On Sep 12, 1996 this sequence version replaced gi:1407373.
 Other_ESTs: TNC83368
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (http://www.tigr.org/tdb/hgi/hgi.html)
 Seq primer: M13-21.

FEATURES

source

Location/Qualifiers
 1..269
 /organism="Homo sapiens"
 /db_xref="ATCC (host):151254"
 /db_xref="taxon:9606"
 /clone_lib="Infant brain"
 /sex="female"
 /dev_stage="infant"
 /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
 Site_2: NotI"

BASE COUNT 54 a 78 c 66 g 71 t

ORIGIN

Query Match 1..18; Score 46; DB 32; Length 269;
 Best Local Similarity 100.0%; Pred. No. 2e-07;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1939
 |||||

Db 61 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 106

RESULT 24

LOCUS

DEFINITION AA955031 343 bp mRNA EST 17-MAR-1999
 Oq6h02.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1591347 3'
 similar to contains Alu repetitive element;contains element MER22

```

repetitive element ;, mRNA sequence.
ACCESSION   AA955031
VERSION     AA955031.1  GI:3118726
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 343)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL     On Jan 19 1998 this sequence version replaced gi:2151892.
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Stratagene, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2046 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 320.
FEATURES
    source
        1..343
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1591347"
            /clone_lib="NCI-CGAP_Kid6"
            /sex="mixed"
            /tissue_type="kidney tumor"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Organ: kidney; Vector: Bluescript SK-; Site:1:
            EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
            Oligo dt. Pooled kidney tumors. 5' adaptor sequence: 5'
            GAATTCGGCAGCAG 3' adaptor sequence: 5'
            CTCAGATTTTTTTTTTTT 3' Average insert size: 1.0 kb."
BASE COUNT  59 a 103 c 88 g 93 t
ORIGIN
Query Match 1.1%; Score 46; DB 40; Length 343;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTTCGAAGTGATTCCTGCTTACGCTCC 1939
|||||
Db 70 CAACCTCTGCTCCCGGGTTCGAAGTGATTCCTGCTTACGCTCC 115

RESULT 25
LOCUS      AQ068513
DEFINITION 10-JUN-1999
            sapiens genomic clone Plate=2124 Col=18 Row=D, genomic survey
            sequence.
ACCESSION  AQ068513
VERSION    AQ068513.1  GI:5068507
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 434)
            Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
            Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
            Hood, L.

```

```

sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
MEDLINE
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2124 row: D column: 18
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 434.
            Location/Qualifiers
                1..434
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Plate=2124 Col=18 Row=D"
                    /clone_lib="CIT Approved Human Genomic Sperm Library D"
                    /sex="male"
                    /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
                    E-Coli DH10B"
BASE COUNT  124 a 111 c 104 g 93 t 2 others
ORIGIN
Query Match 1.1%; Score 46; DB 105; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1886 GGTCACTCAACCTCTGCTCCCGGGTTCGAAGTGATTCCTGCTCC 1931
|||||
Db 252 GGTCACTCAACCTCTGCTCCCGGGTTCGAAGTGATTCCTGCTCC 207

RESULT 26
LOCUS      R50700
DEFINITION yj60g12.r1 Soares breast 2NDH8at Homo sapiens cDNA clone
            IMAGE:153190 5', similar to gb:X59434 THIOSULFATE SULFURTRANSFERASE
            (HUMAN); contains Alu repetitive element; contains MSRI repetitive
            element ;, mRNA sequence.
ACCESSION  R50700
VERSION    R50700.1  GI:812602
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 426)
            Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Travaaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and
            Wilson, R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilton RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 926
            High quality sequence stops: 203 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 926 Std Error: 0.00
            Seq primer: M13Rp1

```



```

RESULT 29
AA654395      180 bp      mRNA      EST      23-DEC-1997
LOCUS
DEFINITION nt03e09.sl NCI_CGAP_Lym3 Homo sapiens cDNA clone IMAGE:1192072 3'
similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION AA654395
VERSION AA654395.1 GI:2590549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 180)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520479.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

Tissue Procurement: Mark Raffeld, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1..180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1192072"
/clone_lib="NCI_CGAP_Lym3"
/tissue_type="lymphoma"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lymph node; Vector: Bluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo df. Pooled lymphomas. 5' adaptor sequence: 5'
GAATCGGCACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 0.9 kb."
BASE COUNT 32 a 52 c 46 g 50 t
ORIGIN
Query Match 1.0%; Score 43; DB 36; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1903 CTCCTCCGGGTTCAAGTGATTCTCTGCTTAGCCTCCCGAGTA 1945
|||||
Db 70 CTCCTCCGGGTTCAAGTGATTCTCTGCTTAGCCTCCCGAGTA 112
|||||

RESULT 30
AA126033      383 bp      mRNA      EST      02-SEP-1999
LOCUS
DEFINITION w01b09.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2457881 3'
similar to contains Alu repetitive element:contains element MER9
repetitive element ;, mRNA sequence.
ACCESSION AA126033
VERSION AA126033.1 GI:5661997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS Hillier.L., Lennon.G., Becker.M., Bonaldo.M.F., Chlapelli.B.,
Chisoe.S., Dietrich.N., Dubuque.T., Favello.A., Gish.W.,
Hawkins.M., Hultman.M., Kucaba.T., Lacy.M., Le.M., Le.N.,
Mardis.E., Moore.M., Morris.M., Parsons.J., Prange.C., Rifkin.L.,
Rohlfing.T., Schellenberg.K., Soares.M.B., Tan.F., Thierry-Mieg.J.,
Trevaskis.E., Underwood.K., Wohlmann.P., Waterston.R., Wilson.R.
and Marra.M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)

```

```

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 383)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137633.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 375.
Location/Qualifiers
1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2457881"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo df.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"
BASE COUNT 85 a 103 c 89 g 105 t 1 others
ORIGIN
Query Match 1.0%; Score 43; DB 62; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1903 CTCCTCCGGGTTCAAGTGATTCTCTGCTTAGCCTCCCGAGTA 1945
|||||
Db 71 CTCCTCCGGGTTCAAGTGATTCTCTGCTTAGCCTCCCGAGTA 113
|||||

RESULT 31
AA121819/c      391 bp      mRNA      EST      02-FEB-1997
LOCUS
DEFINITION z095c09.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone
IMAGE:565984 3' similar to contains Alu repetitive element;contains
element MER22 repetitive element ;, mRNA sequence.
ACCESSION AA121819
VERSION AA121819.1 GI:1679469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS Hillier.L., Lennon.G., Becker.M., Bonaldo.M.F., Chlapelli.B.,
Chisoe.S., Dietrich.N., Dubuque.T., Favello.A., Gish.W.,
Hawkins.M., Hultman.M., Kucaba.T., Lacy.M., Le.M., Le.N.,
Mardis.E., Moore.M., Morris.M., Parsons.J., Prange.C., Rifkin.L.,
Rohlfing.T., Schellenberg.K., Soares.M.B., Tan.F., Thierry-Mieg.J.,
Trevaskis.E., Underwood.K., Wohlmann.P., Waterston.R., Wilson.R.
and Marra.M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)

```

```

MEDLINE
COMMENT
97044478
On May 5, 1995 this sequence version replaced gi:798452.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1369 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 240.
Location/Qualifiers
1. .391
/organism="Homo sapiens"
/db_xref="GDB:4591920"
/db_xref="taxon:9606"
/clone_lib="IMAGE:565984"
/clone_lib="Stratagene fetal retina 937202"
/sex="mixed"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI. Cloned unidirectionally. Primer: Oligo dt. Pooled
retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR
Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCAGTGTTCCTTTTTCCTTTT 3'"
BASE COUNT 103 a 92 c 108 g 78 t 10 others
ORIGIN

Query Match 1.0%; Score 43; DB 28; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1904 CTCCTGGGTTCAAGTATCTCTGCTAGCTCCGAGTAG 1946
|||||
Db 152 CTCCTGGGTTCAAGTATCTCTGCTAGCTCCGAGTAG 110
|||||

RESULT 32
A0515717
LOCUS A0515717 406 bp DNA GSS 05-MAY-1999
DEFINITION HS_5235_A2_E01_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=811 Col=2 Row=I, genomic survey sequence.
ACCESSION A0515717
VERSION A0515717.1 GI:4747975
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 406)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu

FEATURES
source
1. .406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Genomic sequence of BAC ends"
BASE COUNT 77 a 92 c 85 g 148 t 4 others
ORIGIN

Query Match 1.0%; Score 43; DB 104; Length 406;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1898 CTCGTCTCCCGGTTCAAGTATCTCTGCTTACGCTCC 1940
|||||
Db 348 CTCGTCTCCCGGTTCAAGTATCTCTGCTTACGCTCC 390
|||||

RESULT 33
A047553
LOCUS A047553 470 bp mRNA EST 29-SEP-1999
DEFINITION DKFZ586E0121_s1 586 (synonym: hutel) Homo sapiens cDNA clone
DKFZ586E0121, mRNA sequence.
ACCESSION A047553
VERSION A047553.1 GI:4728549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 470)
Ottewaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
EST (Ottewaelder, et al.)
Unpublished (1999)
On Jun 5, 1998 this sequence version replaced gi:3189405.
Contact: Ottewaelder B
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
the German Genome Project.
r1 sequence also available.
This clone is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@r2pd.de.
Location/Qualifiers
1. .470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DKFZ586E0121"
/clone_lib="586 (synonym: hutel)"
/tissue_type="uterus"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI/MluI"
BASE COUNT 125 a 108 c 94 g 143 t
ORIGIN

Query Match 1.0%; Score 43; DB 49; Length 470;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1903 CCTCCCGGTTCAAGTGATTCTCTGCTTAGCTTACGCTCCCGAGTA 1945
|||||
Db 86 CCTCCCGGTTCAAGTGATTCTCTGCTTAGCTTACGCTCCCGAGTA 128

RESULT 34
AQ394316/c AQ394316 527 bp DNA GSS 06-MAR-1999
LOCUS CITBI-E1-2546J9.TF CITBI-E1 Homo sapiens genomic clone 2546J9,
DEFINITION genomic survey sequence.
ACCESSION AQ394316
VERSION AQ394316.1 GI:4365339
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CITBI-E1-2546J9.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
source
location/Qualifiers
1..527
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2546J9"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT 133 a 107 c 135 g 152 t
ORIGIN

Query Match 1.0%; Score 42; DB 102; Length 527;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGTTCAAGTGATTCTCTGCTTAGC 1935
|||||
Db 503 CAACCTCTGCTCCCGGTTCAAGTGATTCTCTGCTTAGC 462

RESULT 35
AQ205998/c AQ205998 389 bp DNA GSS 17-SEP-1998
LOCUS HS_3237_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3237 Col=12 Row=B, genomic survey
sequence.
ACCESSION AQ205998
VERSION AQ205998.1 GI:3616568
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 389)

```

```

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3237 row: B column: 12
Class: BAC ends
High quality sequence stop: 389.
FEATURES
source
location/Qualifiers
1..389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3237 Col=12 Row=B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 123 a 100 c 87 g 77 t 2 others
ORIGIN

Query Match 1.0%; Score 41; DB 99; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1900 CTGCTCCCGGTTCAAGTGATTCTCTGCTTAGCTTACGCTCC 1940
|||||
Db 136 CTGCTCCCGGTTCAAGTGATTCTCTGCTTAGCTTACGCTCC 96

RESULT 36
AQ206027/c AQ206027 400 bp DNA GSS 17-SEP-1998
LOCUS HS_3237_B2_D05_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3237 Col=10 Row=H, genomic survey
sequence.
ACCESSION AQ206027
VERSION AQ206027.1 GI:3616597
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 400)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3237 row: H column: 10
Class: BAC ends
High quality sequence stop: 400.
FEATURES
location/Qualifiers

```

```
source
1..400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3237 Col=10 Row=H"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 135 a 94 c 89 g 82 t
ORIGIN

Query Match 1.0%; Score 41; DB 99; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1900 CTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTCCC 1940
Db 136 CTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTCCC 96

RESULT 37
A0763211/c 513 bp DNA GSS 28-JUL-1999
LOCUS
DEFINITION HS_3222_B2_E12_17C CIT Approved Human Genomic Sperm Library D Homo
sapient genomic clone Plate=3222 Col=24 Row=J, genomic survey
sequence.
ACCESSION A0763211
VERSION A0763211.1 GI:5641327
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 513)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piet@edej.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 3222 row: J column: 24
Seq primer: 17
Class: BAC ends
High quality sequence stop: 513.
Location/Qualifiers
1..513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3222 Col=24 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 166 a 118 c 117 g 110 t 2 others
ORIGIN

Query Match 1.0%; Score 41; DB 87; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1900 CTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTCCC 1940
```

```
Db 154 CTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTCCC 114

RESULT 38
A0742061/c 693 bp DNA GSS 16-JUL-1999
LOCUS
DEFINITION HS_5570_B2_H12_SP6 RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1146 Col=24 Row=P, genomic survey sequence.
ACCESSION A0742061
VERSION A0742061.1 GI:5519583
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piet@edej.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1146 row: P column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 693.
Location/Qualifiers
1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1146 Col=24 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 231 a 131 c 160 g 171 t
ORIGIN

Query Match 1.0%; Score 41; DB 82; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTTAG 1934
Db 494 CAACCTCTGCCTCCGGGTTCAAGTGATTCTCCTAGCCTTAG 454

RESULT 39
A0268072 104 bp DNA GSS 27-APR-1999
LOCUS
DEFINITION RPCI11-7317.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-7317,
genomic survey sequence.
ACCESSION A0268072
VERSION A0268072.1 GI:3795676
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

REFERENCE
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
COMMENT Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1. .104
/organism="Homo sapiens"
/db_xref="GDB:7527846"
/db_xref="taxon:9606"
/clone="RPCI-11-7317"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
20 a 37 c 21 g 26 t

Query Match 0.9%; Score 39; DB 100; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.00013; Mismatches 0; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1902 GCCTCCCGGGTCAAGTGATTCTCCTGCTTAGCCTCCC 1940
|||||
Db 29 GCCTCCCGGGTCAAGTGATTCTCCTGCTTAGCCTCCC 67
|||||
RESULT 40
AA054055 274 bp mRNA EST 01-FEB-1997
LOCUS AA054055
DEFINITION IMAGE:380196 5' similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AA054055
VERSION AA054055.1 GI:1544979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissole,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasaki,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 9704478
COMMENT On May 5, 1995 this sequence version replaced gi:797877.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@atson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2948 Std Error: 0.00
Seq Primer: -28M13 rev2 from AmerSham
High quality sequence stop: 230.
Location/Qualifiers
1. .274

FEATURES
source

/organism="Homo sapiens"
/db_xref="GDB:1288453"
/db_xref="taxon:9606"
/clone="IMAGE:380196"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGCGGCGCGCTTTTITTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."
52 a 79 c 67 g 75 t 1 others
ORIGIN

Query Match 0.9%; Score 39; DB 27; Length 274;
Best Local Similarity 100.0%; Pred. No. 9.2e-05; Mismatches 0; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1902 GCCTCCCGGGTCAAGTGATTCTCCTGCTTAGCCTCCC 1940
|||||
Db 65 GCCTCCCGGGTCAAGTGATTCTCCTGCTTAGCCTCCC 103
|||||

RESULT 41
AQ100978/c 301 bp DNA GSS 27-AUG-1998
LOCUS AQ100978
DEFINITION HS_3061_A1_A03_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3061 Col=5 Row=A, genomic survey sequence.
ACCESSION AQ100978
VERSION AQ100978.1 GI:3472007
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 301)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
 Plate: 3061 row: A column: 5
 Class: BAC ends
 High quality sequence stop: 301.
 Location/Qualifiers

FEATURES

source
 1..301
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="3061 Col=5 Row=A"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOAC11; BAC Clones in E-Coli DH10B"

BASE COUNT

81 a 61 c 79 g 80 t

ORIGIN

Query Match 0.9%; Score 39; DB 91; Length 301;
 Best Local Similarity 100.0%; Pred. No. 8.9e-05;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1902 GCCTCCGGGTCAAGTATCTCGCTTAGCCTCC 1940

Db 299 GCCTCCGGGTCAAGTATCTCGCTTAGCCTCC 261

RESULT 42

AA084176

LOCUS

AA084176 115 bp mRNA EST 31-JUL-1997
 zn17d10.s1 StrataGene neuroepithelium NT2RAM1 937234 Homo sapiens
 cdna clone IMAGE:547699 3' similar to contains Alu repetitive
 element;; mRNA sequence.

ACCESSION

VERSION

AA084176

KEYWORDS

SOURCE

ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 115)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

TITLE

JOURNAL

MEDLINE

COMMENT

On May 8, 1995 this sequence version replaced gi:801257.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1392 Std Error: 0.00
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 97.

FEATURES

source

1..115
 /organism="Homo sapiens"
 /db_xref="GDB:3926215"
 /db_xref="taxon:9606"
 /clone_plate="3061 Col=5 Row=A"
 /clone_lib="StrataGene neuroepithelium NT2RAM1 937234"
 /dev_stage="Ntera-2/RA+MI neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2

(Ntera-2/cl.D1) precursor cells induced with Retinoic
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR
 Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3'
 adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'

BASE COUNT

20 a 36 c 32 g 26 t 1 others

ORIGIN

Query Match 0.9%; Score 38; DB 28; Length 115;

Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCGCCT 1931

Db 35 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCGCCT 72

RESULT 43

H84675/c

LOCUS

H84675 132 bp mRNA EST 14-NOV-1995
 vs66d04.r1 Soares retina N2b4HR Homo sapiens cdna clone
 IMAGE:219751 5' similar to contains Alu repetitive element; contains
 MER22 repetitive element;; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

1..132

/organism="Homo sapiens"

/db_xref="GDB:3847744"

/db_xref="taxon:9606"

/clone_plate="3061 Col=5 Row=A"

/clone_lib="Soares retina N2b4HR"

/sex="male"

/tissue_type="retina"

/dev_stage="55 year old"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: eye; Vector: p773D (Pharmacia) with a
 modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st
 strand cdna was primed with a Not I - oligo(dT) primer [5'
 TTTTACCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTT 3']
 double-stranded cdna was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified p773 vector
 (Pharmacia). The retinas were obtained from a 55 year old
 Caucasian and total cellular poly(A)+ RNA was extracted 6

hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

29 a 34 c 45 g 24 t

Query Match 0.9%; Score 38; DB 24; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 1931
|||||
DB 74 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 37

RESULT 44
AA397419 157 bp mRNA EST 13-APR-1997
LOCUS
DEFINITION
og75d06.s1 NCI_CGAP_Ov8 Homo sapiens cDNA clone IMAGE:1454123 3', similar to contains Alu repetitive element; contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
COMMENT
On Jan 17, 1998 this sequence version replaced gi:2043584.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
source
1..151
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1454123"
/tissue_type="serous adenocarcinoma"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pCMV-SPORT4; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Life Technologies catalog #: 10982-015"

BASE COUNT 25 a 54 c 37 g 34 t 1 others
ORIGIN

Query Match 0.9%; Score 38; DB 40; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 1931
|||||
DB 63 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 100

RESULT 45
AA397419 157 bp mRNA EST 13-AUG-1997
LOCUS
DEFINITION
nc64fll.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:771405 similar to contains Alu repetitive element.;, mRNA Sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
COMMENT
On Apr 14, 1993 this sequence version replaced gi:638990.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Prepared by: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 656 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 155.

FEATURES
source
1..157
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:771405"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors. 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 28 a 50 c 39 g 40 t
ORIGIN

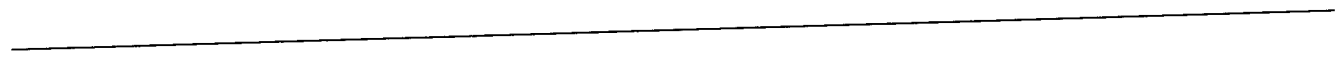
Query Match 0.9%; Score 38; DB 33; Length 157;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 1931
|||||
DB 63 CAACCTCTGCCTCCCGGGTTCAAGTGATTCCTCGCCT 100

Search completed: April 6, 2000, 00:07:41
Job time: 19739 sec

us-09-090-672b-1.oligo_1.rst

Fri Apr 7 09:12:41 2000




```
AC 718551:
DE 06-MAY-1997 (first entry)
DE Human polycystic kidney disease normal PKD1 gene.
KW Adult onset polycystic kidney disease; APKD; autosomal dominant;
KW mutant; transversion; transition; deletion; insertion; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_feature 4379..5272
FT /*tag= a
FT /*note= "specifically claimed region of intronless
FT cDNA identified by exon trapping"
FT old_sequence replace(50652..50653, cg)
FT /*tag= b
FT /*note= "changes Val codon to Leu codon"
FT old_sequence replace(50796..50797, cg)
FT /*tag= c
FT /*note= "replaces Val codon by Leu codon"
FT old_sequence insert(51827..51828, cc)
FT /*tag= d
FT /*note= "insertion, results in frameshift"
FT W09612033-Al.
FT 25-APR-1996.
FT 11-OCT-1995; U13357.
FT 12-OCT-1994; US-323443.
FT 31-JAN-1995; US-381520.
FT (IGIG-) IG LAB INC.
PA (UJO) UNIV JOHNS HOPKINS.
PA Burn TC, Connors TD, Dackowski W, Germino G, Klingler KW;
PI Landes GM, Qian F;
DR WPI; 96-222017/22.
PI Isolated human polycystic kidney disease gene and its mutants
PI useful for treatment of polycystic kidney disease and screening for
PI carriers
PS Claim 1; Fig 1; 65pp; English.
CC The present sequence is that of the normal human PKD1 gene from
CC chromosome 16. Mutations in this gene (e.g. transitions,
CC transversions, deletions and/or insertions) are associated with
CC adult-onset polycystic kidney disease (APKD). The PKD1 locus is
CC GC-rich (62.4%). Comparison of this sequence with a previously
CC reported partial cDNA sequence revealed differences at three
CC locations (see features table). The most significant difference is
CC at position 4566 of the previously reported sequence. The insertion
CC results in a frame-shift in the predicted protein coding sequence,
CC leading to replacement of 92 C-terminal amino acids with a novel
CC 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
CC There is a region consisting of 17 tandem copies of a perfect 27 bp
CC repeat and two large CT-rich regions.
SQ Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T;

Query Match 1.8%; Score 53; DB 1; Length 53577;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 114
|||||
Db 7885 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 7937

RESULT 14
T94108
ID T94108 standard; DNA; 53577 BP.
AC 01-JUN-1998 (first entry)
DE Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.
KW Human; polycystic kidney disease 1; PKD1; treatment;
KW autosomal dominant polycystic kidney disease; APKD; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 27-NOV-1997.
FT W09744457-Al.
FT 22-MAY-1997; U08799.
FT 03-JUN-1996; US-658136.

Query Match 1.8%; Score 53; DB 1; Length 53577;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 114
|||||
Db 7885 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 7937

RESULT 14
T94108
ID T94108 standard; DNA; 53577 BP.
AC 01-JUN-1998 (first entry)
DE Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.
KW Human; polycystic kidney disease 1; PKD1; treatment;
KW autosomal dominant polycystic kidney disease; APKD; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 27-NOV-1997.
FT W09744457-Al.
FT 22-MAY-1997; U08799.
FT 03-JUN-1996; US-658136.

Query Match 1.7%; Score 52; DB 1; Length 1474;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 CACTGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 115
|||||
```

```
PR 24-MAY-1996; US-655360.
PA (GENZ) GENZYME CORP.
PI Burn T, Connors T, Dackowski W, Germino G, Klingler K,
PI Qian F;
DR WPI; 98-018511/02.
PT Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
PS Example 5; Pages 60-89; 257pp; English.
CC The present sequence is the human polycystic kidney disease 1
CC (PKD1) locus between chromosomal markers ATP1 (ATP6C) and D16S84.
CC The PKD1 gene or polypeptide may be used to treat autosomal
CC dominant polycystic kidney disease (APKD), and identify carriers
CC of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC (Ab) that distinguish between normal and mutant PKD1 sequences can
CC also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC perform subcellular and histochemical localisation studies, and to
CC block the function of PKD1. Ab are also useful in rational drug
CC design studies to identify and test inhibitors of PKD1. Sense and
CC antisense sequences derived from the PKD1 gene may be used for
CC detection and therapy.
SQ Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T;

Query Match 1.8%; Score 53; DB 1; Length 53577;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 114
|||||
Db 7884 CTCACGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 7936

RESULT 15
Q12223/c
ID Q12223 standard; DNA; 1474 BP.
AC Q12223;
DT 13-SEP-1991 (first entry)
DE Erbb-3 gene clone E3-1 v-erbB homologous region.
KW Epidermal growth factor; EGFR; receptor; erbB; antibodies; tumour;
KW v-erbB; probe; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 66..221 /*tag= a
FT exon 657..855 /*tag= b
FT exon 973..1446 /*tag= c
FT W09108214-A.
FT 13-JUN-1991.
FT 30-NOV-1990; U07025.
PR 01-DEC-1989; US-444406.
PA (USDC) US SEC OF COMMERCE.
PA (USSH) NAT INST OF HEALTH.
PI Kraus WH, Aaronsen SA;
DR WPI; 91-193144/26.
DR P-PSDB; R12607.
PT Epidermal growth factor receptor related gene - used for
PT producing receptors and antibodies for detecting and treating
PT tumours
PS Disclosure; Fig 3; 34pp; English.
CC The 1.5 kbp region from the EcoRI to the PstI site shows the region
CC of homology in the human erbB-3 gene. The region contains 3 ORFs.
CC See Q12224 for full erbB-3 gene.
SQ Sequence 1474 BP; 375 A; 353 C; 387 G; 359 T;

Query Match 1.7%; Score 52; DB 1; Length 1474;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 CACTGCAAGCTCCGCTCTGGGTTACGCCCAATTCCTGCTCAGCCTCC 115
|||||
```


Db 636 CACTGAAGCTCCGCTCTCTGGTTACGCCATTCTCTGCTCAGCCTCC 585

RESULT 16

ID Q03369 standard; DNA; 1494 BP.
AC Q03369:
DT 20-DEC-1990 (first entry)
DE Sequence encoding human lysozyme
KW Transgenic animal; enzyme; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT signal_peptide 14..67
FT mat_peptide 68..460
FT /*tag= a
FT /*tag= b
PN J02005879-A.
PD 10-JAN-1990.
PF 21-JUN-1988: 151106.
PR 21-JUN-1988: JP-151106.
PA (TAKE) Takeda Chemical Ind KK.
DR WPI: 90-053915/08.
DR P-PSDB: R06108.
PT Human lysozyme gene -
PS useful promoter for genetic expression of animal culturing cell
PS ; Fig 4-1 to 4-3 Pages 834-835; 13pp; Japanese.
CC Human lysozyme DNA can be used to obtain transgenic animals, e.g. mice.
CC It can also be produced by culturing animal cells, and it can
CC promote the genetic expression of these animal cells.
SQ Sequence 1494 BP; 441 A; 298 C; 309 G; 446 T;

Query Match 1.7%; Score 52; DB 1; Length 1494;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGAGGTTTCACC 191
|||||
DB 890 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGAGGTTTCACC 941

RESULT 17

ID N92386 standard; DNA; 1496 BP.
AC N92386:
DT 03-JUN-1990 (first entry)
DE DNA coding for the signal peptide and mature protein of human lysozyme
DE (HL)
KW Human lysozyme; HL; bacterial infection.
OS Homo sapiens.
FH Key Location/Qualifiers
FT signal_peptide 14..67
FT mat_peptide 68..457
FT /*tag= a
FT /*tag= b
FT /*tag= c
PN J01074989-A.
PD 20-MAR-1989.
PF 16-SEP-1987: 229752.
PR 16-SEP-1987: JP-229752.
PA (TAKE) Takeda Chemical Ind KK.
DR WPI: 89-127529/17.
DR P-PSDB: P93510.
PT New DNA sequence -
PT encoding signal peptide and mature protein of human lysozyme
PS Figure 2-1 - 2-1; pages 12-13; 13pp; Japanese.
CC It is useful for effective expression of human lysozyme in yeast or
CC animal cells. Also DNA encoding various proteins can be linked to the
CC 3' end of the HL signal peptide so that it is downstream of the
CC promoter of the expression vector. Expression of such proteins is
CC possible in E. coli, Bacillus subtilis, yeast or animal cells. Lysozyme

Query Match 1.7%; Score 52; DB 1; Length 1494;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGAGGTTTCACC 191
|||||
DB 636 CACTGAAGCTCCGCTCTCTGGTTACGCCATTCTCTGCTCAGCCTCC 585

RESULT 19

V58735/c
ID V58735 standard; cDNA; 1542 BP.
AC V58735;

CC is useful for controlling bacterial infection and unlike chicken lysozyme
CC HL has little adverse effects.
SQ Sequence 1496 BP; 443 A; 299 C; 308 G; 446 T;

Query Match 1.7%; Score 52; DB 1; Length 1496;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 140 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGAGGTTTCACC 191
|||||
DB 890 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGAGGTTTCACC 941

RESULT 18

T10060/c
ID T10060 standard; DNA; 1542 BP.
AC T10060:
DT 13-MAY-1996 (first entry)
DE erbB-3 genomic sequence.
KW erbB-3; antibody; proto-oncogene; cancer; receptor tyrosine kinase;
KW tumour; diagnosis; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 66..221
FT /*tag= a
FT /*codon_start= 66..68
FT Intron 222..779
FT /*tag= b
FT exon 778..855
FT /*tag= c
FT Intron 856..1039
FT /*tag= d
FT exon 1040..1185
FT /*tag= e
PN US5480968-A.
PD 02-JAN-1996.
PF 04-DEC-1989; US-444406.
PR 10-NOV-1992; US-978895.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Aaronson SA, Kraus MH;
DR WPI: 96-068302/07.
DR P-PSDB: R88459.
PT erbB-3 receptor protein binding site peptide and specific antibody -
PT useful in assays to detect the expression of erbB proto-oncogene(s)
PT Example 2: Column 27-30; 41pp; English.
CC A human DNA fragment (T10060) codes for a portion (R88459) of a new
CC functional member of the erbB proto-oncogene family, designated
CC erbB-3. The DNA was identified by screening a human genomic DNA
CC with a probe specific for an oncogenic vital form of the erbB
CC gene, v-erbB, and was subsequently cloned from a human thymus
CC genomic library. The erbB-3 gene was mapped to 12q13. Elevated
CC levels of erbB-3 mRNA were demonstrated in certain mammary tumour cell
CC lines. The gene, or corresponding full-length cDNA (see T10059), can
CC be used to prepare erbB-3 or to detect expression of erbB-3 genes in
CC normal and tumour tissues
SQ Sequence 1542 BP; 389 A; 371 C; 397 G; 385 T;

Query Match 1.7%; Score 52; DB 1; Length 1542;
Best Local Similarity 100.0%; Pred. No. 2.6e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 64 CACTGAAGCTCCGCTCTCTGGTTACGCCATTCTCTGCTCAGCCTCC 115
|||||
DB 636 CACTGAAGCTCCGCTCTCTGGTTACGCCATTCTCTGCTCAGCCTCC 585

Veterinary School, 3800 Spruce Street, Philadelphia, PA 19104-6008,
USA

FEATURES

source

Location/Qualifiers
1. .2253
/organism="Hyalophora cecropia"
/db_xref="taxon:7123"
/tissue_type="fat body"
/dev_stage="5th instar larva"
10. .57
/product="arylphorin"
10. .2124
/note="hemolymph aromatic amino acid rich storage protein;
ArH"

sig_peptide

CDS

/codon_start=1
/product="arylphorin precursor"
/protein_id="AAB86644.1"
/db_xref="GI:2625146"
/translation="MTYLVLAGLIALVQSSVSPVAHHYKTKDVAFTVERQKKVLS
LFQDVQVNDIEYKIGDIDVEANIDNTNKKAVEEFKLYRTGILPKIEFEFLFH
HKREAIALLFLHYADFEFFKSAFAARVHNEGFLYAYIAIQRNTHGFVL
PAYEYPOFFVMOTTVRIYTRKQDGLHTRKAINIGIVKEEHHYVYANYSTFL
YNNDEORTLYTEDIGFNSYIYFHSGLPFWMTSGRYGNLKHRRGEIYFYQQLTR
YFERLNTLGISIPESFWSYPIKTYGTYPLMTSYYPAPQENYLNHVKYNAIRFL
DIEKTFVSLQKGFESNGKIDPFDEKALNFYQWQENADLYGEEYTKDYQSYSE
IVARHLVGAAPKPFDKHIFMPSALDFYOTALRDPFQYQNLRIYGVINAKHYLVYP
KSLRVNIEVKVSDVANDVLIKIPMGPKYDDNGFPITLLENWKKFFELDNTHKLTGQN
KIVRNNEFLIKEDSLPLDLLLDEKGVFPFDMSEFGIMPKRLMLPRGTEGIPF
QFFVYIPSESSHDLTPFEAPVINDKPLGYPFDPDPVDSFSKQPNFFKDVVTVHEG
EHFPYINVPYSIKTNEVPKQ"

58. .2121

/note="N-terminal sequence of the mature protein in the
blood"

/evidence="experimental"

/product="arylphorin"

BASE COUNT 717 a 482 C 414 G 640 T

ORIGIN

Query Match 1.4%; Score 32; DB 35; Length 2253;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2245 TTTATTAAAAA...AAAAAAAAAAAAAAAAA 2276
|||||...|||||...|||||...|||||...
Db 2202 TTTATTAAAAA...AAAAAAAAAAAAAAAAA 2233
|||||...|||||...|||||...|||||...
RESULT 23
AF032398 2431 bp mRNA INV 28-SEP-1999
LOCUS
DEFINITION
Hyalophora cecropia moderately methionine rich hexamerin precursor,
mRNA, complete cds.
ACCESSION
AF032398
VERSION
AF032398.1 GI:2625149
KEYWORDS
cecropia moth.
SOURCE
Hyalophora cecropia
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
Ditrysia; Bombycoidea; Saturniidae; Saturniinae; Hyalophora.
1 (bases 1 to 2431)
AUTHORS
Massey, H.C. Jr.
TITLE
The Evolution of Insect Storage Proteins: Driven by Composition and
Constrained by Sequence
JOURNAL
Thesis (1995) University of Pennsylvania
REFERENCE
2 (bases 1 to 2431)
AUTHORS
Burmester, J., Massey, H.C. Jr., Zakharkin, S.O. and Benes, H.
TITLE
The evolution of hexamerins and the phylogeny of insects
JOURNAL
J. Mol. Evol. 47 (1), 93-108 (1998)
MEDLINE
98329402
REFERENCE
3 (bases 1 to 2431)
AUTHORS
Massey, H.C. Jr. and Telfer, W.H.

TITLE

JOURNAL

Direct Submission
Submitted (03-NOV-1997) Pathobiology, University of Pennsylvania
Veterinary School, 3800 Spruce Street, Philadelphia, PA 19104-6008,
USA

FEATURES

source

Location/Qualifiers
1. .2431
/organism="Hyalophora cecropia"
/db_xref="taxon:7123"
/tissue_type="fat body"
/dev_stage="5th instar larva"
32. .82
/product="moderately methionine rich hexamerin"
32. .2290
/note="moderately methionine rich fat body storage
protein; Mth"

sig_peptide

CDS

/codon_start=1
/product="moderately methionine rich hexamerin precursor"
/protein_id="AAB86646.1"
/db_xref="GI:2625150"
/translation="MRAAVLLVGLAAIASARPNDNDVNFVSMIDKQQLFLKLLNH
VTEPYKDVSEIGKDFEEDNMELTKTEVKSFMVLMKVGLPRGEVFTLVNDROM
KEVNNHILYIAKDFSTFKTACWMLYLNEMFVYLSVAVRRECKGVLPDPY
EYPPYFRAVDIOKALIKMRKGLDQLKLSFEYIKKTKEDVYIIDENVYDRVALN
EDDKYFTEIDIDNTYVYFHVDPFWMKDQIEDKLTFRRELTYIQOILARVYL
FELNGEIEPESFYPIQGYWPKLHNGIEUPARENKVSXYKNTDELISVE
YETLKEALMGYIDMGIRLELTKEPEDTETLGMKLYGKAAKQDSKCEVYKILLI
VKSAGLUNTSKDYFVAPTVLDSTOTALRDPFYQLOKRLNVIVQLFKRLPCYTK
BELYPPGYKLVNVDKLLTYFDYFDMNTNAVTLSEEEKKASMSGFVKKRLNH
EFKYSBELSKAVDCVIRIFMGPKEDQDLRIDINKRLNFVEMDSFVYKLTGKN
TIVKSSDMHNLVPRMNTSDMKWVESITMDRLDFVKDLKNTYFTGFTLLPKGSV
SLQPMVTVVYSPKLVDNVDMSALDPTSKALSFEDSTVLMKMPGLGFLDRYIDV
TNFTSNKFDYDVTIFHKNSQCDMKRWKDLKSYNLASKTPIESNTYFVDTDLNS
SKVSSANVDM"

83. .2287

/note="alternate form of the mature protein in the blood"

/evidence="experimental"

/product="moderately methionine rich hexamerin"

86. .2287

/note="alternate form of the mature protein in the blood"

/evidence="experimental"

/product="moderately methionine rich hexamerin"

BASE COUNT 777 a 474 C 487 G 693 T

ORIGIN

Query Match 1.4%; Score 32; DB 35; Length 2431;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2245 TTTATTAAAAA...AAAAAAAAAAAAAAAAA 2276
|||||...|||||...|||||...|||||...
Db 2399 TTTATTAAAAA...AAAAAAAAAAAAAAAAA 2430
|||||...|||||...|||||...|||||...
RESULT 24
DDICNPA
LOCUS
DEFINITION
Dictyosellum discoideum cyclic nucleotide phosphodiesterase gene,
complete cds.
ACCESSION
M23449 M34483
VERSION
M23449.1 GI:695312
KEYWORDS
cyclic nucleotide phosphodiesterase.
SOURCE
D.discoideum (strain NC4; cell line AX3-K) DNA, clones pGP-1 and
pC34.
ORGANISM
Dictyostelium discoideum
Eukaryota; Dictyostellida; Dictyostelium.
1 (bases 1 to 6372)
REFERENCE
Podgoriski, G.J., Franke, J., Faure, M. and Kessin, R.H.
TITLE
The cyclic nucleotide phosphodiesterase gene of Dictyostelium
discoideum utilizes alternate promoters and splicing for the
synthesis of multiple mRNAs
JOURNAL
Mol. Cell. Biol. 9 (9), 3938-3950 (1989)
MEDLINE
89384622

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 09:00:38 ; Search time 345.99 Seconds
(without alignments)
930.476 Million cell updates/sec

Title: US-09-090-672B-2
Perfect score: 2689
Sequence: 1 GTTGGAGGTTCTGGGGCGCA.....AAAAAAAAAAAAAAAAAAAA 2689

Scoring table: OLIGO_NUC

scoring cable: CUI900_RSC
Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : C

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Minimum DB seq	length:	0
Maximum DB seq	length:	1000000

post-processing: Listing first 45 summaries

```
Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata1/ina/6_COMB.seq.*
6: /cgn2_6/ptodata1/ina/PCtUS9_COMB.s
7: /cgn2_6/ptodata1/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	1.7	2989	7	5378464-1	Patent No. 5378464-1
2	45	1.7	3350	3	US-08-663-566A-1	Sequence 1, Appli
3	45	1.7	3350	4	US-08-023-610-1	Sequence 1, Appli
4	45	1.7	3350	4	US-08-288-065A-1	Sequence 1, Appli
5	45	1.7	3350	4	US-08-362-240A-1	Sequence 1, Appli
6	45	1.7	3350	6	PCR-US95-1024S-1	Sequence 1, Appli
7	44	1.6	1391	1	US-08-281-652-1	Sequence 1, Appli
8	44	1.6	1391	6	PCR-US95-0775S-1	Sequence 1, Appli
9	44	1.6	1781	1	US-08-334-615-2	Sequence 2, Appli
10	44	1.6	1781	1	US-08-334-362-2	Sequence 2, Appli
11	44	1.6	1781	1	US-08-433-010-2	Sequence 2, Appli
12	44	1.6	1781	2	US-08-245-295-4	Sequence 4, Appli
13	44	1.6	1781	2	US-08-481-130-4	Sequence 4, Appli
14	44	1.6	1781	2	US-08-656-984A-4	Sequence 4, Appli
15	44	1.6	1781	2	US-08-482-882-2	Sequence 2, Appli
16	44	1.6	1781	2	US-08-483-604-4	Sequence 2, Appli
17	44	1.6	1781	2	US-08-483-389-2	Sequence 2, Appli
18	44	1.6	1781	3	US-08-487-1130-2	Sequence 2, Appli
19	44	1.6	1781	3	US-08-487-595-4	Sequence 4, Appli
20	44	1.6	1781	3	US-08-473-503-2	Sequence 2, Appli
21	44	1.6	1781	3	US-08-483-932-2	Sequence 2, Appli
22	44	1.6	1781	4	US-08-720-420A-2	Sequence 2, Appli
23	44	1.6	2158	2	US-08-698-551-1	Sequence 1, Appli
24	44	1.6	2158	3	US-08-603-228-1	Sequence 1, Appli
25	44	1.6	2158	3	US-08-649-341A-1	Sequence 1, Appli
26	44	1.6	2158	3	US-08-494-440B-1	Sequence 1, Appli

27	44	1.6	2158	3	US-08-533-901B-1	Sequence 1, Appl 1
28	44	1.6	2158	3	US-08-839-032A-1	Sequence 1, Appl 1
29	44	1.6	2158	4	US-08-839-032A-1	Sequence 1, Appl 1
30	44	1.6	2158	6	PT-US93-12724-1	Sequence 2, Appl 1
31	44	1.6	3471	6	PT-US93-10227-2	Sequence 2, Appl 1
32	43	1.6	75	4	US-08-776-944-13	Sequence 13, Appl 1
33	43	1.6	80	2	US-07-920-281C-25	Sequence 25, Appl 1
34	43	1.6	635	2	US-08-455-632A-35	Sequence 35, Appl 1
35	43	1.6	635	2	US-08-416-336-5	Sequence 35, Appl 1
36	43	1.6	635	4	US-08-456-460C-35	Sequence 35, Appl 1
37	43	1.6	635	6	PT-US94-05352A-35	Sequence 35, Appl 1
38	43	1.6	688	7	5498694-3	Patent No. 5498694
39	43	1.6	1013	1	US-07-920-519-30	Sequence 30, Appl 1
40	43	1.6	1013	1	US-08-086-410-23	Sequence 23, Appl 1
41	43	1.6	1013	1	US-08-314-586-30	Sequence 30, Appl 1
42	43	1.6	1386	4	US-08-897-340-1	Sequence 1, Appl 1
43	43	1.6	1493	1	US-08-340-820-24	Sequence 24, Appl 1
44	43	1.6	1493	1	US-08-593-525-24	Sequence 24, Appl 1
45	43	1.6	1512	4	US-08-909-9650-8	Sequence 8, Appl 1

ALIGNMENTS

RESULT 1
5378464-1
; Patent NO. 5378464
; APPLICANT: MCIVER, RODGER P.
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/320,408
; FILING DATE: 08-MAR-1989
; SEQ ID NO:1:
; LENGTH: 2989
5378464-1

```
Query Match      1.7%; Score 45; DB 7; Length 2989;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 45; Conservative 0; Mismatches 0; Indels
```

QY	2645	CCCCA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2689
Db	2911	CCCCA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2955

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```


Medical Research Council
Addenbrookes Hospital, Hills Road, Cambridge, CB2 2QH, England
Fax: 44 (0) 1223 412178
Email: phd@mc-lmb.cam.ac.uk

Primer A: GAGTACCTAAGCTTACTGACAGG
Primer B: TCTTAGTCGGGAAGAGAGG
STS size: 95

Protocol:
Template: 10 - 50 ng
Primer: 1 uM each
DNTPs: 200 uM each
Taq Pol: 0.025 units/uL
Total Vol: 10 uL

Buffer:
MgCl2: 1 mM
KCl: 50mM
Tris HCl: 10 mM
pH: 8.3

Alu-PCR product from mouse/human hybrid cell line 10479 (Coriell Cell Repositories).

FEATURES

source
1. .277
/organism="Homo sapiens"
/db_xref="taxon:9606"

STS

primer_bind

BASE COUNT 81 a 61 c 59 g 76 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 27; DB 13; Length 277;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1416 TGGGCAACAGAGCAAGACCCAGTCTCA 1442

Db 18 TGGGCAACAGAGCAAGACCCAGTCTCA 44

RESULT 30

HUMUT154

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 23, 1995 this sequence version replaced gi:464065.

Submitted by: Utah Center for Human Genome Research University of

Utah, Dept. of Human Genetics

2160 Eccles Institute of Human Genetics

Salt Lake City, UT 84112

e-mail: sts@corona.med.utah.edu

Primer A: ACTACACTCCAGTCTGGGCAACA

Primer B: GTAGCTCCCTCTCTTTCCCTGCT

PCR Profile:

Initial Denaturation: 94C 0sec

Cycles Denaturation Annealing Extension 30 94
C 60 sec. 60 C 60 sec. 72 C 60 sec. Mg++: 1.50 mM Gel:
Acrylamide 7%, Formamide 32%, Urea 34%
Alleles: 1.
FEATURES Location/Qualifiers
source 1. .370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17"
primer_bind 8. .30
/evidence=experimental
STS 8. .134
/standard_name="STS UT154, GDB D17S615"
primer_bind complement(112..134)
/evidence=experimental
BASE COUNT 105 a 77 c 104 g 80 t 4 others
ORIGIN
Query Match 1.8%; Score 27; DB 13; Length 370;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1416 TGGGCAACAGAGCAAGACCCAGTCTCA 1442
Db 22 TGGGCAACAGAGCAAGACCCAGTCTCA 48
RESULT 31
G07947/c 31
LOCUS G07947 471 bp DNA STS 05-FEB-1997
DEFINITION human STS CHLC.CTT16.P18204 clone CTT16, sequence tagged site.
ACCESSION G07947
VERSION G07947.1 GI:938497
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human vector-pUCPI host-E.coli dut+ung+ (DH10B) Marker Selected genomic DNA prepared from XY individual of French nationality.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.
TITLE Cooperative Human Linkage Center
JOURNAL Unpublished (1995)
COMMENT Synonyms: CTT16, CHLC.CTT16.#79987
Contact: Dr. Jeffrey C. Murray
UofI
The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jeff-murray@uiowa.edu
Primer A: TGGCTAGTGGACAAAGTGST
Primer B: GCCTGTATTACTGTGTTCTT
STS size: 174
PCR Profile:
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C
Protocol:
Template: 30ng genomic DNA
Primer: each 1.5 pmole
DNTPs: each 200 uM
Taq Polymerase: 0.5 units
Total Vol: 10 uL
Buffer:
MgCl2: 1.5mM
KCl: 50mM
Tris: 10mM

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
12140200
On Dec 3, 1996 this sequence version replaced gi:1119186.
Other ESTs: THC126123
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 274
/organism="Homo sapiens"
/db_xref="ATCC (inhost):135902"
/db_xref="taxon:9606"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 87 a 51 c 53 g 81 t 2 others
ORIGIN
Query Match 27.2%; Score 74.6; DB 32; Length 274;
Best Local Similarity 98.7%; Pred. No. 7.4e-08;
Matches 74; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 49 TGCCTATATACCAACTGATAAATCTAGATTCTGTCGTCTATAGAC 108
Db 1 TGCCTATATACCAACTGATAAATCTAGATTCTGTCGTCTATAGAC 60
QY 109 CATGTTGTAGG 123
Db 61 CATGTTGTAGG 75

RESULT 11
LOCUS AA488906 559 bp mRNA EST 15-AUG-1997
DEFINITION aa5a02.r1 NCI-CCAP-GCB1 Homo sapiens cDNA clone IMAGE:824810 5' similar to TR:G607003 G607003 BETA TRANSDUCIN-LIKE PROTEIN. ;, mRNA sequence.
ACCESSION AA488906
VERSION AA488906.1 GI:2218508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 559)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397871.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 446.
Location/Qualifiers
1. 559
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CCAP-GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - Oligo(dT) primer [5'-TGTTACCAATCAAGTGGAGCGCCCTCATTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 158 a 139 c 100 g 162 t
ORIGIN

Query Match 23.6%; Score 64.6; DB 34; Length 559;
Best Local Similarity 98.5%; Pred. No. 1.3e-05;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 59 ACAATACCAACTGATAAATCTAGATTCTGTCGTCTATAGACCATGTTGTA 118
Db 559 ACAATACCAACTGATAAATCTAGATTCTGTCGTCTATAGACCATGTTGTA 500
QY 119 GTAGG 123
Db 499 GTAGG 495

RESULT 12
LOCUS T10350 259 bp mRNA EST 13-APR-1994
DEFINITION seg1091 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.
ACCESSION T10350
VERSION T10350.1 GI:471699
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 259)
AUTHORS Soares, M.B., Bonaldo, M.F., Jelenc, P., Su, L., Lawton, L. and Efstathiadis, A.
TITLE Construction and characterization of a normalized cDNA library
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)
MEDLINE 95023884
COMMENT Contact: Bento Soares
Columbia University
Department of Psychiatry, 722 W 168th Street, Unit #41, New York, NY 10032


```

exon      2288..2678
          /gene="rp15"
          /citation=[2]
          /number=4
          /evidence=experimental
3' UTR    2519..2678
          /gene="rp15"
          /citation=[2]
          /evidence=experimental
polyA_site 2678..2680
          /citation=[2]
          /evidence=experimental

BASE COUNT 920 a 522 c 550 g 992 t
ORIGIN
Query Match 9.1%, Score 18; DB 7; Length 2984;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 94 AAAATATATATATGCTG 111
    |||||
DB 798 AAAATATATATATGCTG 781

RESULT 35
AF102623 7186 bp DNA BCT 24-MAY-1999
DEFINITION Methanosarcina barkeri dimethylamine corrinoid protein MtbC (mtbc),
            trimethylamine methyltransferase MtbB (mtbb), trimethylamine
            corrinoid protein MtbC (mttc), putative transmembrane protein MtbP
            (mtbp), and dimethylamine methyltransferase MtbB1 (mtbb1) genes,
            complete cds.
ACCESSION AF102623
VERSION AF102623.1 GI:4262423
KEYWORDS
SOURCE Methanosarcina barkeri.
ORGANISM Methanosarcina barkeri.
            Methanosarcinae; Methanosarcinaceae;
            Methanosarcina;
            Methanosarcina barkeri.
REFERENCE Ferguson,D.J., Gorlatova,N., Paul,L., Grahame,D. and Krzycki,J.A.
AUTHORS Ferguson,D.J., Jr. and Krzycki,J.A.
TITLE Reconstitution of trimethylamine-dependent coenzyme M methylation
            with the trimethylamine corrinoid protein and the isozymes of
            methyltransferase II from Methanosarcina barkeri
JOURNAL J. Bacteriol. 179 (3), 846-852 (1997)
MEDLINE 97158682
REFERENCE
AUTHORS Ferguson,D.J., Gorlatova,N., Paul,L., Grahame,D. and Krzycki,J.A.
TITLE The corrinoid protein from Methanosarcina barkeri specific for
            dimethylamine: COM methyl transfer
JOURNAL Unpublished
REFERENCE
AUTHORS Paul,L. and Krzycki,J.A.
TITLE The genes encoding the trimethylamine and dimethylamine
            methyltransferases of Methanosarcina barkeri are cotranscribed and
            interrupted by translationally bypassed in-frame amber codons
            Unpublished
JOURNAL 4 (bases 1 to 7186)
AUTHORS Paul,L. and Krzycki,J.A.
TITLE Direct Submission
SUBMITTER Submitted (27-OCT-1998) Microbiology, Ohio State University, 484
            West 12th Ave, Columbus, OH 43210, USA
LOCATION/Qualifiers
1. 7186
/organism="Methanosarcina barkeri"
/strain="MS"
/db_xref="taxon:2208"
734..1375
/gene="mtbc"
734..1375
/gene="mtbc"
/function="participates in dimethylamine:coenzyme M methyl

```

```

transfer"
/note="ORF identified by N-terminal sequence of isolated
corrinoid protein"
/codon_start=1
/transl_table=11
/product="dimethylamine corrinoid protein MtbC"
/protein_id="ABD14632.1"
/db_xref="GI:4262424"
/transl_table="MSKELLQELADATISCKRDTVAVAVKANGELPSELTIERKLA
AGNEVGVFERGKLFPLHVMADADMTAGVALLDLPEDGASGSKGLVINGVEGD
VHDGKAIVSTIOTASGEVHDIGRDVPIRNFIERKENVNDMISALMTTLIOGSK
SVLELKEGELDKKVKVVGAPATQAMADKIGADCYENNTAEVAKAKELLA"
1402..2889
/gene="mtb"
1402..2889
/gene="mtb"
1402..2889
/gene="mtb"
/note="UG codon is translated by an undetermined
mechanism; ORF identified with N-terminal sequence of 52
kDa trimethylamine methyltransferase polypeptide"
/codon_start=1
/transl_table="pos:2401..2403;aa:OTHER"
/transl_table=11
/product="trimethylamine methyltransferase MtbB"
/evidence=experimental
/transl_table="MANKNAVAGFNAINGVELNLTFTDELRNHYANMEVINDRGIOV
SDEARQITRENGCEVNEKTVKIPETLVRKQLAPREFTLMGRDRTYQDECG
KRWTFGTVGVVCKVYODKRTVYDSEKRIADALCWMANIDRLSPARSADNG
OGADVHEHLEPLANTANFHHIDPVEGNEYRDIVRKYVYGGDEEKRPISMLL
CPSPELSELVNACQVYIIGARFGIPVAVLSMAAGSSSVYIAGTLVHNKVEVLSGLV
LADLTPGAKVWVGSSSTTTPDLKKGTPVGSPELGISAAYAKLAQFYGNPSVYAGX
SDAKVPDQAGHEKMTTLPLALAGANTIGAGMELGWTFSMEQVLDINDIFSVMKX
AMGIVPSEETLSEVIOKVGIGNNFPALQZQLVDVYSNPMILDRHMEFGDMAAGS
KDLATVAHEKVEDVLEKNHQTVPIDADIFKQMAQVYDADKARGM"
2920..3573
/gene="mttc"
2920..3573
/gene="mttc"
/function="participates in trimethylamine:coenzyme M
methyl transfer with TMA methyltransferase"
/note="ORF identified by N-terminal sequence of corrinoid
protein"
/codon_start=1
/transl_table=11
/product="trimethylamine corrinoid protein MtbC"
/protein_id="ABD14631.1"
/db_xref="GI:4262425"
/transl_table="MANKETIIRAKKATIDPDELAEVANEALAGIDPEVLEIKG
FTGMEVEEKRGQGLLPLPVLAARAMNSGKIVITPEMKRSQTRSGTVAIGTI
EISIHISGDIIVASMLNLAGFRVVDLDGVDPIITFEKVEKELKPOVANSALMTTIV
NQDIEQKLEAGVDOVKTYVGGAPVTOADMADKIGADIYGSANDAVAKKALNV"
3885..4946
/gene="mtbp"
3885..4946
/gene="mtbp"
/function="possible methylamine permease"
/codon_start=1
/transl_table=11
/product="putative transmembrane protein MtbP"
/protein_id="ABD14632.1"
/db_xref="GI:4262427"
/transl_table="MEGVFVDLAKLAKQESKRIISKYMAFECAIFNGIYLBGTIV
WVNPFDMSAIAKTSIGDGLSVTVVLTITANALAVMLAVMNVGLKRGELVET
LKFEHPCSKWFLASIFGPAALIGSFVANGFSAVADGLPVVAGSLAAYWG
EKISRAAIGIIVYIGISYGGGLFELSSGNVPMVIGGLMAAGMGIEGALNG
KGLDIAEPNAGLTLREGENIIMWIIIVPLALVGPMSFALQADPDLTILIVFG
ITFGPCYVWYKSPPLIGVGRGIGNLVGLACIHLFLFPEDVDWITIIIGALCIA
GSPMTEPDSLETLRGE"
5470..6873
/gene="mtbb1"
5470..6873
/gene="mtbb1"
/function="

```

4

6

8



BASE COUNT	174 a	123 c	96 g	203 t			
ORIGIN							
Query Match	23.2%	Score 63.6;	DB 104;	Length 596;			
Best Local Similarity	94.3%;	Pred. No. 2.2e-05;					
Matches 66:	Conservative	0; Mismatches 4;	Indels	0; Gaps 0;			
QY	1	CCTTTACAGATTCTTCGCGCTCGGGTGAACACTACAAGGATCGGTGGCCTATATCAC	60				
		() () () () () () () () () ()					
Db	527	CGTTTACAGACTCTTCGCGCAGCGGTGGAACACTACAAGGATCGGTGGCCTATATCAC	586				
		() () () () () () () () () ()					
QY	61	AATACCAAC 70					
		() () () () () () () () () ()					
Db	587	AATACCAAC 596					
RESULT 14							
LOCUS	AI196903	449 bp	mRNA	EST 14-OCT-1998			
DEFINITION	ui55c03.yl Sugano mouse liver mlia Mus musculus cDNA clone IMAGE:1886308 5' similar to TR:O14432 O14432 TUP1.; mRNA sequence.						
ACCESSION	AI196903						
VERSION	AI196903						
KEYWORDS	EST.						
SOURCE	house mouse.						
ORGANISM	Mus musculus						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
AUTHORS	1 (bases 1 to 449) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T., Geisel,S., Kuback,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.						
TITLE	The washU-HHMI Mouse EST Project						
JOURNAL	Unpublished (1996)						
COMMENT	Contact: Marra M/Mouse EST Project washU-HHMI Mouse EST Project Washington University School of Medicine# 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:970632 Seq primer: custom primer used High quality sequence stop: 448.						
FEATURES	Location/Qualifiers						
source	1..449						
	/organism="Mus musculus"						
	/strain="C57BL"						
	/db_xref="taxon:10090"						
	/clone="IMAGE:1886308"						
	/clone_lib="Sugano mouse liver mlia"						
	/sex="female"						
	/dev_stage="adult"						
	/lab_host="DH10B"						
	/note="Organ: liver; Vector: pME18S-FL3; Site_1: DralIII (CACATG); Site_2: DraII (CACATG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACATGCTG, 3' site CACATGCTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCCTGCTCTAAAGCTGG and 3' end primer CGACCTGCAGCTCGACACA."						

PA (GENZ) GENZYME CORP.
PI Burn T, Connors T, Dackowski W, Germino G, Klinger K,
PI Qian F;
DR WPI: 98-018511/02.
PT Human polycystic kidney disease gene, PKD1 - useful to treat and
PT diagnose human autosomal or adult onset polycystic kidney disease
PS Example 5; Pages 60-89; 237pp; English.
CC The present sequence is the human polycystic kidney disease 1
CC (PKD1) locus between chromosomal markers ATP6C and D16S84.
CC The PKD1 gene or polypeptide may be used to treat autosomal
CC dominant polycystic kidney disease (APKD), and identify carriers
CC of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies
CC (Ab) that distinguish between normal and mutant PKD1 sequences can
CC also be used in diagnostic tests. Anti-PKD1 Ab may also be used to
CC perform subcellular and histochemical localisation studies, and to
CC block the function of PKD1. Ab are also useful in rational drug
CC design studies to identify and test inhibitors of PKD1. Sense and
CC antisense sequences derived from the PKD1 gene may be used for
CC detection and therapy. 8495 A; 17684 C; 15782 G; 11616 T;
SQ Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T;
Query Match 1.4%; Score 45; DB 1; Length 53577;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3110 GAGACGAGCTGACCAACATGCTGAACCCGCTCTACTATAAAA 3154
DB 1692 GAGACGAGCTGACCAACATGCTGAACCCGCTCTACTATAAAA 1648
RESULT 13
ID X02711 standard; DNA: 1083 BP.
AC X02711:
DT 14-MAY-1999 (first entry)
DE Human transaldolase TAL-H TARE-6 retrotransposable element.
KW Transaldolase; TAL-H; autoantigen; human; antibody; neuroprotector;
KW neurodegenerative autoimmune disease; multiple sclerosis; detection;
KW proliferation inhibitor; cytotoxicity inhibitor; binding inhibitor;
KW retrotransposon; retrotransposable element; TARE-6; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_signal 538..929
FT /*tag= a
FT /note= "Alu-sc-like dimer"
FT US5879909-A.
PN 09-MAR-1999.
PD 09-OCT-1998; 057762.
PR 19-OCT-1994; US-326119.
PR 09-APR-1998; US-057762.
PA (UUNY) UNIV NEW YORK STATE RES FOUND.
PI Perl A;
DR WPI: 99-203948/17.
PT Isolated human transaldolase gene - useful for raising antibodies
PT for detecting neurodegenerative autoimmune diseases, especially
PT multiple sclerosis
PS Claim 12; Fig 19; 55pp; English.
CC This sequence represents a human transaldolase (TAL-H) retrotransposable
CC element TARE 6. The TAL-H protein can be used in methods to raise
CC antibodies for detecting human transaldolase-mediated neurodegenerative
CC autoimmune diseases, especially multiple sclerosis. The protein has
CC neuroprotective properties and is a proliferation, cytotoxicity and
CC binding inhibitor.
SQ Sequence 1083 BP; 237 A; 302 C; 252 G; 292 T;
Query Match 1.3%; Score 44; DB 1; Length 1083;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2566 GGGGTTTCACTGTTAGCCAGGATGCTTGATCTCTCTGACCT 2609
|||||

DB 822 GGGGTTTCACTGTTAGCCAGGATGCTTGATCTCTCTGACCT 865
RESULT 14
V57903/c
ID V57903 standard; DNA: 237326 BP.
AC V57903:
DT 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
PN W09814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
PI WPI: 98-240014/21.
DR Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 1; Fig 9; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BTF genes, which are homologues of the milk
CC protein butyrophilin (BT), and can be used in the production of agonists
CC and antagonists of BT function. Also described are: (1) a RoRet gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia.
SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;
Query Match 1.3%; Score 44; DB 1; Length 237326;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3111 AGACCAGCTGACCAACATGCTGAACCCGCTCTCTACTATAAAA 3154
DB 49851 AGACCAGCTGACCAACATGCTGAACCCGCTCTCTACTATAAAA 49808
RESULT 15
Q62613/c
ID Q62613 standard; DNA: 1688 BP.
AC Q62613:
DT 17-JAN-1995 (first entry)
DE Human mdr-1 promoter fragment.
KW Inducible promoter; cytostatic agent; adriamycin; vincristine;
KW multiple drug resistance; mdr-1; mammalian expression vector;
KW cancer therapy; ss.
OS Homo sapiens.
PN DE4238778-A.
PD 19-MAY-1994.
PF 12-NOV-1992; 238778.
PR 12-NOV-1992; DE-238778.
PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Stein U, Walther W;
DR WPI: 94-168680/21.
PT New mammalian expression vector useful for gene therapy, -
PT comprising the promoter and/or the enhancer of the mdr-1 gene
PT which is susceptible to cytostatic agents.
PS Claim 2; Page 5; 7pp; German.
CC This is a preferred fragment of the mdr-1 gene promoter, isolated
CC from human DNA. The mdr-1 promoter and enhancer elements are
CC inducible by cytostatic agents such as vincristine and adriamycin
CC which are used in cancer therapy. Vectors comprising the promoter
CC and enhancer sequences operably linked to heterologous genes coding
CC for therapeutic agents are claimed. Using the vectors, the
CC therapeutic agents (e.g. anti-oncogenes, interferon, IL-2, IL-6,
CC etc.) can be expressed in cancer cells under the control of
CC cytostatic agents.
SQ Sequence 1688 BP; 556 A; 311 C; 371 G; 450 T;

Query Match 1.3%; Score 43; DB 1; Length 1688;
Best Local Similarity 100.0%; Pred. No. 3.5e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2567 GGGTTTCACGTGTAGCCAGGATGCTTGTATCTCTGACCT 2609
|||||
DB 272 GGGTTTCACGTGTAGCCAGGATGCTTGTATCTCTGACCT 230
|||||

RESULT 16

ID Q62612 standard; DNA; 2090 BP.
AC Q62612;
DT 17-JAN-1995 (first entry)
DE Human mdr-1 promoter fragment.
KW Inducible promoter; cytostatic agent; adriamycin; vincristine;
KW multiple drug resistance; mdr-1; mammalian expression vector;
KW cancer therapy; ss.
OS Homo sapiens.
PN DE4238778-A.
PD 19-MAY-1994.
PF 12-NOV-1992; 238778.
PR 12-NOV-1992; DB-238778.
PA (DELB-) DELBRUCK CENT MOLEKULARE MEDIZIN MAX.
PI Stein U, Walther W;
DR WPI: 94-168680/21.
PT New mammalian expression vector useful for gene therapy, -
PT comprising the promoter and/or the enhancer of the mdr-1 gene
PT which is susceptible to cytostatic agents.

PS Claim 2; Page 4; 7pp; German.
CC This is a preferred fragment of the mdr-1 gene promoter, isolated
CC from human DNA. The mdr-1 promoter and enhancer elements are
CC inducible by cytostatic agents such as vincristine and adriamycin
CC which are used in cancer therapy. Vectors comprising the promoter
CC and enhancer sequences operably linked to heterologous genes coding
CC for therapeutic agents are claimed. Using the vectors, the
CC therapeutic agents (e.g. anti-oncogenes, interferon, IL-2, IL-6,
CC etc.) can be expressed in cancer cells under the control of
CC cytostatic agents.
SQ Sequence 2090 BP; 714 A; 365 C; 444 G; 567 T;

Query Match 1.3%; Score 43; DB 1; Length 2090;
Best Local Similarity 100.0%; Pred. No. 3.5e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2567 GGGTTTCACGTGTAGCCAGGATGCTTGTATCTCTGACCT 2609
|||||
DB 674 GGGTTTCACGTGTAGCCAGGATGCTTGTATCTCTGACCT 632
|||||

RESULT 17

ID Q81139 standard; DNA; 15328 BP.
AC Q81139;
DT 17-JAN-1995 (first entry)
DE Human brain expressed sequence tag ESR02702.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 440; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes

DT 15-AUG-1995 (first entry)
DE HPLA2-8 gene.
KW HPLA2-8; phospholipase A2; PLA2; Batten disease;
KW neuronal ceroid lipofuscinosis; gene therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 399..466
FT /*tag= a
FT /*label= Exon-1
FT 2633..2777
FT /*tag= b
FT /*label= Exon-2
FT 13862..14028
FT /*tag= c
FT /*label= Exon-3
PN WO9502328-A.
PD 26-JAN-1995.
PF 15-JUL-1994; U07926.
PR 15-JUL-1993; US-091941.
PR 26-JUL-1993; US-097354.
PA (INCY-) INCYTE PHARM INC.
PA (INDV) UNIV INDIANA FOUND.
PI Seilhamer JJ, Tischfield JA;
DR WPI: 95-067096/09.
DR P-PSDB: R63052; R63053.
PT Novel type III and IV low mol. wt. phospholipase A2 enzymes -
PT from humans and rats, also nucleic acid sequences useful, e.g.
PT for recombinant prodn. of enzymes, research into Batten's
PT disease, etc.
PS Claim 8; Page 67-76; 160pp; English.
CC A human PLA2-encoding cDNA (Q81138) expressing HPLA2-10, was
CC isolated from human brain RNA by RACE-PCR. 2 rat PLA2 cDNAs,
CC designated RPLA2-8 (Q81136) and RPLA2-10 (Q81137), were isolated from
CC rat brain and heart cDNA libraries, respectively. A partial human
CC genomic counterpart to RPLA2-8, RPLA2-8 (Q81139), was also obtained.
CC RPLA2-8 and HPLA2-8 have been designated type III PLA2, and RPLA2-10
CC and HPLA2-10 as type IV. 3885 A; 3789 C; 4082 G; 3572 T;
SQ Sequence 15328 BP; 3885 A; 3789 C; 4082 G; 3572 T;

Query Match 1.3%; Score 43; DB 1; Length 15328;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3231 TGAACCCGGGAGCGAGTTGTCAGTGCAGCGGAGATCACACCA 3273
|||||
DB 1790 TGAACCCGGGAGCGAGTTGTCAGTGCAGCGGAGATCACACCA 1832
|||||

RESULT 18

ID Q61108 standard; DNA; 223 BP.
AC Q61108;
DT 16-MAR-1994 (first entry)
DE Human brain expressed sequence tag ESR02702.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 440; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes

27782. 27861,29276. 29347,29456. 29570,29695. 29742.
 30178. 30207,30365. 30473,33547. 33605,34826. 34955,
 35525. 35643))
 /gene="SAPK4"
 /note="match: proteins 014739 015124"
 /codon_start=1
 /evidence="not_experimental"
 /product="dj179N16.1 (Stress Activated Protein Kinase 4
 (Mitogen Activated Protein Kinase p38delta))"
 /protein_id="CAB08438.1"
 /db_xref="GI:3036774"
 /db_xref="GI:3036774"
 /db_xref="SPTREMBL:O15264"
 /translation="MSLRKKFKYQDNKTNWELPKTYVSPHYGSGAYGVSQAID
 KRSGEVAIKLSRPFQSEIFAKRAYRELLKMHQENHVGILDDVTFPSSLRNFYD
 FYLVNPFQDLOLRKMGSESEKIQYLVYQMLKGLKIHSGVHRLDKPGNLAVNE
 DCELKILDFGLARHADAEMTYVYVTRWRAPEVILSMYHNTQVDSVSGVMAEMLT

Query Match 13.9% Score 19; DB 10; Length 172048;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCACCGCACCTGGCTGATG 19
 |||||
 b 41760 CCACCGCACCTGGCTGATG 41742

RESULT 18
 LOCUS AC007158 204027 bp DNA HTG 26-OCT-1999
 DEFINITION Homo sapiens clone hRPK.90_A_1, *** SEQUENCING IN PROGRESS ***, 2
 ordered pieces.
 AC007158
 AC007158
 AC007158.5 GI:5757580
 HTG; HTSG; PHASE2.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 204027)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArrellano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Melidrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
 Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 22, 1999 this sequence version replaced gi:4726136.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.

* the accession number will be preserved
 * 1 160375: contig of 160375 bp in length
 * 160376 204027: contig of 43652 bp in length.
 FEATURES
 source
 1. 204027
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hrpk.90_A_1"
 /clone_lib="RPGI-11 Human Male BAC"
 BASE COUNT 63998 a 40107 c 35047 g 60875 t
 ORIGIN

Query Match 13.9% Score 19; DB 42; Length 204027;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGATG 19
 |||||
 DB 111376 CCACCGCACCTGGCTGATG 111394

RESULT 19
 LOCUS HUMVIPHMI 270 bp DNA PRI 14-JAN-1995
 DEFINITION Human vasoactive intestinal peptide and histidine-methionine amide
 peptide hormone (VIP/PHM-27) gene, exon 1.
 M14618
 M14618.1 GI:340266
 VERSION histidine-methionine amide; hormone; peptide hormone; vasoactive
 KEYWORDS intestinal peptide.
 SEGMENT 1 of 6
 SOURCE Human DNA (libraries of Lawn and EMBL), clones lambda-VIP-D and
 lambda-VIP-4.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 270)
 Linder,S., Barkhem,T., Norberg,A., Persson,H., Schalling,M.,
 Hokfelt,T. and Magnusson,G.
 TITLE Structure and expression of the gene encoding the vasoactive
 intestinal peptide precursor
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (2), 605-609 (1987)
 MEDLINE 87092456
 COMMENT Draft entry and clean copy of sequence [1] kindly provided by
 S.Linder, 02-MAR-1987.
 FEATURES
 source
 1. 270
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="6q24-q27"
 prim_transcript 100. >270
 /gene="VIP"
 /note="VIP/PHM27 mRNA: G00-120-490"
 264. >270
 /gene="VIP"
 /note="VIP/PHM27 mRNA intron A"
 BASE COUNT 76 a 67 c 69 g 58 t
 ORIGIN Unreported.

Query Match 13.1% Score 18; DB 9; Length 270;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TGACTTCTTTCAGAGGAC 47
 |||||
 Db 22 TGACTTCTTTCAGAGGAC 39

RESULT 20

PS Example: Fig 2A-H; 98pp; English.
 CC A cDNA molecule (T63334) codes for human nucleoprotein interacto
 CC 1 (NPI-1) (W16327), a host cell protein which interacts with
 CC influenza virus nucleoprotein (NP) and which may be an accessory
 CC protein required for influenza virus replication. It is the human
 CC homologue of the yeast protein SRP1. NPI-1 was identified by
 CC interactive trap selection using LexA-NP as bait and yeast
 CC transformed with a HeLa cDNA library, and by 5' RACE (see also
 CC T63341). NPI-1 to 6 cDNAs (see also T63335-39) have been
 CC cloned and analysed. The human NS1-1 gene (T63340), which codes
 CC for a protein (W16329) that interacts with influenza virus protein
 CC NS1 was similarly identified. The cDNA clones can be used to
 CC produce human host cell proteins for use in screening for cpds.
 CC that inhibit viral replication.
 SQ Sequence 2940 BP; 826 A; 641 C; 627 G; 846 T;

Query Match 11.9%; Score 16; DB 1; Length 2940;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ACAGAGATTAGACAAA 108
 |||||
 DB 766 ACAGAGATTAGACAAA 751

RESULT 6
 X36424
 ID X36424 standard; DNA: 536 BP.
 AC X36424;
 DT 06-JUL-1999 (first entry)
 DE Human heretulin-like factor coding sequence clone HAGF38R.
 DE Human heretulin-like factor; HLF; cell growth regulator; diagnosis;
 KW human heretulin-like factor; HLF; cell growth regulator; diagnosis;
 OS Homo sapiens
 PN W09857989-A1.
 PD 23-DEC-1998. U12403.
 PF 16-JUN-1998; US049942.
 PR 17-JUN-1997; US049942.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (GEOU) UNIV GEORGETOWN.
 PI Hijazi MM, King CR, Ruben SM, Young P;
 DR WPI: 99-095327/08.
 PT New isolated heretulin-like factor - used to develop products for
 PT the diagnosis and treatment of disorders involving regulation of
 PT cell growth, particularly cancers
 PS Claim 20; Page 90-91; 118pp; English.
 CC This sequence encodes the human heretulin-like factor (HLF) of the
 CC invention. The HLF is involved in the regulation of cell growth.
 CC Detection of different levels of expression of the HLF gene can be used
 CC for the diagnosis of disorders, e.g. in the neural system. In
 CC particular, detection of different levels of HLF gene expression in cells
 CC or body fluid of an individual can be used for diagnosing cancer. The
 CC products can also be used in the treatment of disorders involving
 CC abnormal levels of HLF activity.
 SQ Sequence 536 BP; 141 A; 131 C; 125 G; 120 T;

Query Match 11.1%; Score 15; DB 1; Length 536;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGGGCTTCCCTTT 39
 |||||
 DB 472 AGGGCTTCCCTTT 486

RESULT 7
 N80469
 ID N80469 standard; cDNA: 713 BP.
 AC N80469;
 DT 06-DEC-1990 (first entry)
 DE Gastrin inhibitory polypeptide precursor.

KW Gastrin inhibitory polypeptide precursor; GIP; probe; insulin;
 KW diabetes; ss.
 OS Homo sapiens.
 FH Key
 CDS 101...562
 Location/Qualifiers
 FT 101...562
 /tag= a
 FT /label=GIP precursor
 FT /note="Claim 1"
 signal_peptide 1...160
 /tag= b
 FT 254...379
 /tag= c
 FT /product=processed GIP
 FT 290...303
 /tag= d
 FT /label=probe
 FT /note="used to design probe for diabetes diagnosis"
 other_feature 356...372
 /tag= e
 FT /label=probe
 FT /note="used to design probe for diabetes diagnosis"
 poly_a_signal 653...658
 /tag= f
 FT 692...697
 /tag= g
 FT /note="most likely of the two signals"
 BP-269072-A.
 01-JUN-1988.
 PD 24-NOV-1987; 117325.
 PR 27-NOV-1986; JP-282812.
 PA (SANW) SANWA KAGAKU KENKYUSHO.
 PI Takeda J, Imura H, Seino Y, Tanaka K, Takahashi H, Mitani T;
 PI Kurono M, Sawai K;
 DR WPI: 88-148897/22.
 DR P-PSDB: P80287.
 PT DNA encoding human gastric inhibitory polypeptide precursor -
 PT used as probe for diagnosis of diabetes and for producing
 PT polypeptide(s) for diabetes treatment.
 PS Disclosure: P; English.
 CC The sequence was obtd. from a clone isolated from a cDNA library
 CC prepared from total RNA extracted from the duodenum of a patient
 CC during panceato-duodenectomy for pancreatic cancer. The sequence
 CC can be used to design probes (see features) for the diagnosis of
 CC diabetes. It may also be ligated into an expression vector for
 CC prodn. of recombinant GIP. GIP accelerates the gastric secretion
 CC and insulin secretion and can be used to treat diabetes.
 CC See also N82432 and N82433.
 SQ Sequence 713 BP; 193 A; 187 C; 191 G; 142 T;

Query Match 11.1%; Score 15; DB 1; Length 713;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGAAGAAAGAGGTC 30
 |||||
 DB 165 AGAAGAAAGAGGTC 179

RESULT 8
 V15706
 ID V15706 standard; cDNA: 863 BP.
 AC V15706;
 DT 03-JUL-1998 (first entry)
 DE Recombinant mosquito salivary allergen rAed a 3 cDNA.
 DE Recombinant mosquito; salivary allergen; rAed a 3; determination;
 KW bite sensitivity; epi-cutaneous test; skin test; intradermal test;
 KW allergy diagnosis; immunotherapy; desensitisation; ss.
 OS Aedes aegypti.
 FH Key
 CDS 12...773
 /tag= a
 FT 12...773
 FT sig_peptide 12...65

Example 2," fig 8; 209pp: English

The present invention describes hereditary haemochromatosis gene products from the human haemochromatosis gene. The present sequence represents a hereditary haemochromatosis subregion from an individual unaffected by hereditary haemochromatosis (HH). Also described is a method to determine the presence or absence of the common hereditary haemochromatosis (HFE) gene mutation in an individual comprising: (a) providing DNA or RNA from the individual; and (b) assessing the DNA or RNA for the presence or absence of a haplotype or genotype where the presence or absence of the haplotype genotype indicates the likely presence of the HFE gene mutation in the genome of the individual. The HFE gene sequences from the present invention can be used to develop products for use in the diagnosis and treatment of HFE. The present invention also describes Bfr genes, which are homologues of the milk protein butyrophilin (BT), and can be used in the production of agonists and antagonists of BT function. Also described are: (1) a Roret gene which can be used to develop products for the study, diagnosis and treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes which are homologues of a type 1 sodium transport gene, and can similarly be used for hypophosphatemia.

RESULT	6
Q60985	
ID	Q60985 standard; DNA; 382 Bp.
AC	Q60985;
DT	16-MAR-1994 (first entry)
DE	Human brain Expressed Sequence Tag ESO1661.


```

KEYWORDS      HTG.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          1 (bases 1 to 32070).
JOURNAL        Bentley/D. and Maggall, L.
REFERENCE      The sequence of H. sapiens cosmid LUC422
AUTHORS        2 (bases 1 to 32070)
TITLE          Unpublished (1996)
JOURNAL        Waterston/R.
COMMENT        Direct Submission
                Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO 63110, USA, and
                e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
: all regions were double stranded or sequenced with an alternate
  chemistry; an attempt was made to resolve all sequencing problems,
  such as compressions and repeats; all regions were covered by
  sequence from more than one subclone; and the assembly was
  confirmed by restriction digest

SOURCE INFORMATION:
This clone is from a chromosome 3 specific library, VECTOR: pWE15
Clone reference: Ming-Hui Wei et al, CANCER RESEARCH,
56,1487-1492,1996.

NEIGHBORING SEQUENCE INFORMATION:

The left clone is LUC421: the right clone is 3938P1. Actual start
of this cosmid is at base position 1 of HUMDUC422.
Location/Qualifiers
1. 32070
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="3"
   /clone="LUC422"
   /map="3p21.3"
   /map="3p21.3"
1. 235
   /rpt_family="ALU"
728. 1009
   /rpt_family="ALU"
   complement(1876..2168)
   /rpt_family="ALU"
   complement(10132..10417)
   /rpt_family="ALU"
10458. 10603
   /rpt_family="ALU"
   complement(10601..10786)
   /rpt_family="ALU"
   complement(11435..29883)
   /gene="LUC415"
CDS
   complement(join(11435..11560,12513..12642,12721..12818,
13912..13986,14285..14464,15641..15712,15793..15942,
16348..16509,19112..19200,19419..19503,20194..20279,
21578..21650,21735..21812,22312..22398,23017..23114,
24169..24329,24736..24801,25570..25630,26712..26795,
29269..29342,29823..29883))
   /gene="LUC415"
   /note="H_LUC422.1: coded for by human cDNAs U23946,

repeat_region .12379)
rpt_family="ALU"
complement(14318..14464)
/gene="LUC415"
/note="Grail prediction, score = 88"
repeat_region .15245)
rpt_family="ALU"
complement(14715..15245)
/gene="LUC415"
/note="Grail prediction, score = 82"
misc_feature .16739)
/gene="LUC415"
/note="Grail prediction, score = 82"
misc_feature .16739)
/gene="LUC415"
/note="match: human 3' EST N74644 bases 1-340"
16400..16673)
/gene="LUC415"
/note="16739"
16673..16739)
/gene="LUC415"
/note="match: human 5' EST W00982 bases 283-343"
16740..17031)
rpt_family="ALU"
complement(17457..17759)
rpt_family="ALU"
18209..18449)
rpt_family="ALU"
18664..18737)
/note="LUC415"
/note="match: human 3' EST N59570 bases 355-428"
complement(19663..19951)
rpt_family="ALU"
20510..20800)
rpt_family="ALU"
complement(21166..21455)
rpt_family="ALU"
21898..22188)
rpt_family="ALU"
23353..23915)
rpt_family="ALU"
complement(24579..24684)
rpt_family="ALU"
24685..25296)
/gene="LUC415"
/note="close match: multiple ESTs, see AA001762 and
R01204"
complement(25227..25519)
rpt_family="ALU"
complement(26223..26512)
rpt_family="ALU"
27411..27832)
/gene="LUC415"
/note="close match: human 3' EST AA037589 and 5' EST
AA03758"
28604..28733)
misc_feature

```

Fri Apr 7 09:13:44 2000

us-09-090-672b-2.oligo.rge

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:50:51 ; Search time 9481.22 Seconds

(without alignments)
-861.159 Million cell updates/sec

Title: US-09-090-672B-2

Perfect score: 2689

Sequence: 1 GTTGAGGCTCTGGGCGCA.....AAAAAAAAAAAAAAAAAAAA 2689

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pl3: *
10: gb_pl4: *
11: gb_pl5: *
12: gb_pl6: *
13: gb_pl7: *
14: gb_pl8: *
15: gb_pl9: *
16: gb_pl10: *
17: gb_pl11: *
18: gb_pl12: *
19: gb_pl13: *
20: gb_pl14: *
21: gb_pl15: *
22: gb_pl16: *
23: gb_pl17: *
24: gb_pl18: *
25: gb_pl19: *
26: gb_pl20: *
27: gb_pl21: *
28: gb_pl22: *
29: gb_pl23: *
30: gb_pl24: *
31: gb_pl25: *
32: gb_pl26: *
33: gb_pl27: *
34: gb_pl28: *
35: gb_pl29: *
36: gb_pl30: *
37: gb_pl31: *
38: gb_pl32: *
39: gb_pl33: *
40: gb_pl34: *
41: gb_pl35: *
42: gb_pl36: *
43: gb_pl37: *

44: gb_pl38: *
45: gb_pl39: *
46: gb_pl40: *
47: gb_pl41: *
48: gb_pl42: *
49: gb_pl43: *
50: gb_pl44: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	57.5	32070	10 HUMLUCA22	U73168 Human cosmi
2	518	19.3	3135	40 AF091263	AF091263 Homo sapi
3	467	17.4	3094	40 AF103802	AF103802 Homo sapi
4	379	14.1	2575	10 HSU23946	U23946 Human puta
5	168	6.2	83030	40 AC004814	AC004814 Homo sapi
6	120	4.5	132	5 A74798	A74798 Sequence 48
7	120	4.5	132	5 A77777	A77777 Sequence 48
8	47	1.7	113	35 SC096099	U96099 Sarcophaga
9	47	1.7	2805	12 RATLUNAT	W7345 Rat L-type
10	47	1.7	184553	43 AC013452	AC013452 Homo sapi
11	46	1.7	1305	5 A69673	A69673 Sequence 3
12	46	1.7	1306	5 A69681	A69681 Sequence 11
13	46	1.7	1651	11 AF067420	AF067420 Homo sapi
14	46	1.7	3832	10 HSU75370	U75370 Human mito
15	46	1.7	12971	35 AF088979	AF088979 Dictyoste
16	46	1.7	28862	41 AC010238	AC010238 Homo sapi
17	46	1.7	78250	44 AC016250	AC016250 Homo sapi
18	46	1.7	107475	44 AC017106	AC017106 Homo sapi
19	46	1.7	108064	41 AC010404	AC010404 Homo sapi
20	46	1.7	110000	45 AC017014	AC017014 Homo sapi
21	46	1.7	132333	41 AC009889	AC009889 Homo sapi
22	46	1.7	140513	33 AL133259	AL133259 Homo sapi
23	46	1.7	144387	33 HSDJ315G1	AL121601 Homo sapi
24	46	1.7	163195	43 AC010129	AC010129 Homo sapi
25	46	1.7	163533	43 AC011280	AC011280 Homo sapi
26	46	1.7	165284	42 AC012057	AC012057 Homo sapi
27	46	1.7	165414	42 AC012457	AC012457 Homo sapi
28	46	1.7	165445	42 AC009969	AC009969 Homo sapi
29	46	1.7	168363	42 AC012114	AC012114 Homo sapi
30	46	1.7	169168	44 AC013321	AC013321 Homo sapi
31	45	1.7	342	7 SSY10781	Y10781 S.stapfianu
32	45	1.7	1016	8 AF005158	AF005158 Arabidops
33	45	1.7	1180	10 AF012536	AF012536 Homo sapi
34	45	1.7	3105	10 HSM800649	AF012536 Homo sapi
35	45	1.7	3220	40 AF106934	AF106934 Homo sapi
36	45	1.7	3350	5 AR068182	AR068182 Homo sapi
37	45	1.7	3582	3 S73375	S73375 P25/35-Cyc
38	45	1.7	41684	41 AC008696	AC008696 Homo sapi
39	45	1.7	55868	33 AC008251	AC008251 Drosophila
40	45	1.7	111079	11 HS419C19	AL035407 Human DNA
41	45	1.7	125260	41 AC011414	AC011414 Homo sapi
42	45	1.7	128208	33 HSDJ319W7	AL079341 Homo sapi
43	45	1.7	145346	43 AC011262	AC011262 Homo sapi
44	45	1.7	156730	45 AC015500	AC015500 Homo sapi
45	45	1.7	162504	42 AC009820	AC009820 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS HUMLUCA22/c
DEFINITION Human cosmid LUCA22, complete sequence.
ACCESSION U73168
VERSION U73168.1 GI:161899

DNA
PRT
07-FEB-1997

```
/gene="LUCA15"
/note="close match; human EST Z20369 bases 1-129"
29212..29547
/misc_feature
/gene="LUCA15"
/note="close match; rat 3' EST H33623"
complement(29990..30410)
/rpt_family="ALU"
repeat_region
complement(30609..30901)
/rpt_family="ALU"
repeat_region
30996..31595
/rpt_family="ALU"
repeat_region
complement(31762..32048)
/rpt_family="ALU"
repeat_region
BASE COUNT 8573 a 8148 c 7445 g 7904 t
ORIGIN
Query Match 57.5%; Score 1547; DB 10; Length 32070;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2197; Conservative 0; Mismatches 1; Indels 7; Gaps 3;
QY 446 ATTCAGAAATGATGAGCTTCCTCAAGCCCTCAGCCTCGGATGCGAGCTGATGAAG 505
DB 29892 ATTCAGAAATGATGAGCTTCCTCAAGCCCTCAGCCTCGGATGCGAGCTGATGAAG 29833
QY 506 AGGAAACAGGTGAGAGCTTGCTTAGTTCCTGATATATTGTTCTCTCCCATTCGCCAC 565
DB 29832 AGGAAACAGGTGAGAGCTTGCTTAGTTCCTGATATATTGTTCTCTCCCATTCGCCAC 29773
QY 566 CTCAGTCCCTAAAGAACATCTGATTCCTCCAGCTTCAAGCACATGAATTCAGAAAGAA 625
DB 29772 CTCAGTCCCTAAAGAACATCTGATTCCTCCAGCTTCAAGCACATGAATTCAGAAAGAA 29713
QY 626 AGGTTGGCATGGCTAAGGAATGTCACATCTTGAAGAACATGTTAGCATCTGAGGAAGCTT 685
DB 29712 AGGTTGGCATGGCTAAGGAATGTCACATCTTGAAGAACATGTTAGCATCTGAGGAAGCTT 29653
QY 686 TTTTAAACTTTTGTAGGAGCTTTTTCCTTAGTAAAGTAATGATTTATAACTCCT 745
DB 29652 TTTTAAACTTTTGTAGGAGCTTTTTCCTTAGTAAAGTAATGATTTATAACTCCT - 29593
QY 746 TTTTATTTTGTGACTATAGTCGGTTGCAATGGTTACTTTAAGCGTGGAAATCAATGGAGT 805
DB 29593 -TTTATTTTGTGACTATAGTCGGTTGCAATGGTTACTTTAAGCGTGGAAATCAATGGAGT 29535
QY 806 GGCATTTAGTTCAGCGGCTTGCTTCCCTGCCATGCAAGTATCAAGAGATCCCCAGT 865
DB 29534 GGCATTTAGTTCAGCGGCTTGCTTCCCTGCCATGCAAGTATCAAGAGATCCCCAGT 29475
QY 866 CAAGTCACATTTGTAAGCTGCTTCCCAATTTGCTTCAGCGAGTGTGGAAGCAGTGG 925
DB 29474 CAAGTCACATTTGTAAGCTGCTTCCCAATTTGCTTCAGCGAGTGTGGAAGCAGTGG 29415
QY 926 GAGAGAGATTACCTGTTTATAAGGAACATGACTACACAAGTATCCCGTCTATATCTGAA 985
DB 29414 GAGAGAGATTACCTGTTTATAAGGAACATGACTACACAAGTATCCCGTCTATATCTGAA 29355
QY 986 TGCTGTCTAGTGTGTAAGCCGTGTTTCGCTTCCCTGCGAGTTTATCATTGCAAGATG 1045
DB 29354 TGCTGTCTAGTGTGTAAGCCGTGTTTCGCTTCCCTGCGAGTTTATCATTGCAAGATG 29295
QY 1046 CTACCAAGCTGGATGGAAGCAATCAGTTGCTTCACCTACCAAGTCTAGATATTCATGAA 1105
DB 29294 CTACCAAGCTGGATGGAAGCAATCAGTTGCTTCACCTACCAAGTCTAGATATTCATGAA 29235
QY 1106 AATGGAACAGTCTGTACAAATTTTAAAAAAGGTTGAAGAGTGGTTGTTCCCAAGGAG 1165
DB 29234 AATGGAACAGTCTGTACAAATTTTAAAAAAGGTTGAAGAGTGGTTGTTCCCAAGGAG 29175
QY 1166 TGACTTTTTTTAAAAAAGGCTTTGTTATATATATAAATTTGATCTACTAGATAAGT 1225
DB 29174 TGACTTTTTTTTAAAAAAGGCTTTGTTATATATATAAATTTGATCTACTAGATAAGT 29116
```

```
QY 1226 ACAGTACCAAGGACTTTCATATAGAAATTTGTTCTGCCCTTTAAACATGGCTACCTACCTGG 1285
DB 29115 ACAGTACCAAGGACTTTCATATAGAAATTTGTTCTGCCCTTTAAACATGGCTACCTACCTGG 29056
QY 1286 CAGGGCTTTGTTAACTACTGAATACCTGCTCGTAACTCACTAAACATCTTAAATGTTTCC 1345
DB 29055 CAGGGCTTTGTTAACTACTGAATACCTGCTCGTAACTCACTAAACATCTTAAATGTTTCC 28996
QY 1346 CTTTTTCTAGTTTGTATATATCTTATATTCCTATATGTCCTATTCAGAGTAAGCTTAGTATACAA 1405
DB 29995 CTTTTTCTAGTTTGTATATATTCCTATATGTCCTATTCAGAGTAAGCTTAGTATACAA 28936
QY 1406 CTCCTCATTTGACAGTGAAGAGAACATAGTGAAGTCTGTGGCGGCATTTTATAGTAA 1465
DB 28935 CTCCTCATTTGACAGTGAAGAGAACATAGTGAAGTCTGTGGCGGCATTTTATAGTAA 28876
QY 1466 TTCCTTATTTCTGCTGGAAGACACAAAGCCCTCGGAGCGCTAACTGCTCAGACCGGTC 1525
DB 28875 TTCCTTATTTCTGCTGGAAGACACAAAGCCCTCGGAGCGCTAACTGCTCAGACCGGTC 28816
QY 1526 TTCAGGGAATATTTAAGGACTTAGTGGAAATTTATGAACAATAAGTCTGATGAGATTAGCC 1585
DB 28815 TTCAGGGAATATTTAAGGACTTAGTGGAAATTTATGAACAATAAGTCTGATGAGATTAGCC 28756
QY 1586 TGGAGTGGTCTCGACGCTGCTTAATCTAGTTAGAGTGGCAATTAACATTTCTAATCTCC 1645
DB 28755 TGGAGTGGTCTCGACGCTGCTTAATC---TAGAGTGGCAATTAACATTTCTAATCTCC 28700
QY 1646 TTCAGAAATGCTTTTATAGTCTGTTCAAGCAAGTCATGATGGTTCTTCGAGGTAGTGT 1705
DB 28699 TTCAGAAATGCTTTTATAGTCTGTTCAAGCAAGTCATGATGGTTCTTCGAGGTAGTGT 28640
QY 1706 TAACTGAAAGTGTCTTCTCAGTTTGTCAAGATAATGTTAGTGTGTCAGTCTTGAATAACAT 1765
DB 28639 TAACTGAAAGTGTCTTCTCAGTTTGTCAAGATAATGTTAGTGTGTCAGTCTTGAATAACAT 28580
QY 1766 TTTTTCGAGAAGTCCCAAGGCACATTTATGAATGCCCTTTAAACAAGTGCATTTCTGGGAAG 1825
DB 28579 TTTTTCGAGAAGTCCCAAGGCACATTTATGAATGCCCTTTAAACAAGTGCATTTCTGGGAAG 28520
QY 1826 TTTGCTTGACTCATATTCTTGCTTTTCTCGACATTTCTGTGATTTGAGTCATCCATGAAT 1885
DB 28519 TTTGCTTGACTCATATTCTTGCTTTTCTCGACATTTCTGTGATTTGAGTCATCCATGAAT 28460
QY 1886 CCATGAATAAAGCTTACATTTCTTGATTTGTTATATTTGCAATTTATAACAAGACTCACTA 1945
DB 28459 CCATGAATAAAGCTTACATTTCTTGATTTGTTATATTTGCAATTTATAACAAGACTCACTA 28400
QY 1946 ATGAGGTATCATCTTGTGACTGATTTGTTAAAGTTTTTTAAGCTCTCATTTTCTCTAA 2005
DB 28399 ATGAGGTATCATCTTGTGACTGATTTGTTAAAGTTTTTTAAGCTCTCATTTTCTCTAA 28340
QY 2006 CCCAGAATCACAGCTGATTTTATTTAAAGTAGAGCTTCATTTCTCATACCATAGAT 2065
DB 28339 CCCAGAATCACAGCTGATTTTATTTAAAGTAGAGCTTCATTTCTCATACCATAGAT 28280
QY 2066 ACCATCTAGTAAATCCAGAACATATACAAAGTTTCATGTGAGTCTGCTTTCTTGACATGA 2125
DB 28279 ACCATCTAGTAAATCCAGAACATATACAAAGTTTCATGTGAGTCTGCTTTCTTGACATGA 28220
QY 2126 TAGCATTTGTTGATGCGAGTGGATATGTCAGAAATGACTAACCTTAGGAGTTTAAACCTCTTA 2185
DB 28219 TAGCATTTGTTGATGCGAGTGGATATGTCAGAAATGACTAACCTTAGGAGTTTAAACCTCTTA 28160
QY 2186 AGAACTTAAACCTGTAAAGACATTTAAAGTCTCCACATTTTAAATGTATACAAAGCTAT 2245
DB 28159 AGAACTTAAACCTGTAAAGACATTTAAAGTCTCCACATTTTAAATGTATACAAAGCTAT 28100
QY 2246 GTTACTGTAAACACATTACAGTTCAAATTCACCTCCAGAAATAAAAAGGCCAGTAGGATTA 2305
DB 28099 GTTACTGTAAACACATTACAGTTCAAATTCACCTCCAGAAATAAAAAGGCCAGTAGGATTA 28040
QY 2306 GGGACTCACTGGTAGTTGGAGTCTCCACACACATCCCTCTCTCTAGTGGATGATCTATT 2365
```

```
|||||
Db 28039 GGGACTCACTGGTAGTTGGAGTCTCCAGACACATCCCTCTCTAGTGGATGATCTATT 27980
Qy 2366 CACATATCTCCAGCTTTTATTTTCTCTCTATATCACAGTGAAGTGGGCCCTT 2425
|||||
Db 27979 CACATATCTCCAGCTTTTATTTTCTCTCTATATCACAGTGAAGTGGGCCCTT 27920
Qy 2426 CAGCTTTTCTCTCTGCGCCAGACATCGAGTCTTTCCTTTAGATATCGCAGACAAAAAT 2485
|||||
Db 27919 CAGCTTTTCTCTCTGCGCCAGACATCGAGTCTTTCCTTTAGATATCGCAGACAAAAAT 27860
Qy 2486 TCACAGCATGCTTAAATCTTCCAGGATTTCAGACACCAAAATGCTCACAGTATGTAT 2545
Db 27859 TCACAGCATGCTTAAATCTTCCAGGATTTCAGACACCAAAATGCTCACAGTATGTAT 27800
Qy 2546 GTTTAGAGGGTTAGACTCTCTTTTAAATCTGGATATCTAACACCTACTTAAATCTGT 2605
|||||
Db 27799 GTTTAGAGGGTTAGACTCTCTTTTAAATCTGGATATCTAACACCTACTTAAATCTGT 27740
Qy 2606 TTGATAGTGTCAAAACCCACCCCTTGATCTTCCACCCCCCAA 2650
|||||
Db 27739 TTGATAGTGTCAAAACCCACCCCTTGATCTTCCACCCCCCAA 27695

RESULT 2
AF091263
LOCUS AF091263 3135 bp mRNA PRI 06-OCT-1999
DEFINITION Homo sapiens RNA binding motif protein 5 (RBM5) mRNA, complete cds.
ACCESSION AF091263
VERSION AF091263.1 GI:4140646
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Elst,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
An evolutionary rearrangement of the Xp11.3-11.23 region in 3p21.3,
a region frequently deleted in a variety of cancers
Genomics 60 (2), 238-240 (1999)
99417685
MEDLINE
REFERENCE
2 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Elst,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
A comparison of genomic structures and expression patterns of two
closely related flanking genes in a critical lung cancer region at
3p21.3
Unpublished
JOURNAL
3 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Elst,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
Direct Submission
JOURNAL
Submitted (13-SEP-1998) Medical Genetics, University Groningen,
A.Deusinglaan 4, Groningen 9713AW, The Netherlands
FEATURES
source
1. .3135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/notes="lung cancer region"
1. .3135
/gene="RBM5"
149..2596
/gene="RBM5"
/codon_start=1
/product="RNA binding motif protein 5"
/protein_id="AAD04159.1"
/db_xref="GI:4140647"
/translation="MGSDKRVSRTERSGRYSIIIDRRDRERSRRRSDSDYKRSSD
```

```
DRGRDYDDYDSDPERERERNSRSEDYHSDGYGEHDIYRHDISDERESKTI ML
RGLPITTESDI REMSEFEGPOADVRLMKRTKTVSRGFAVEFYHLDATSWMEAN
QKLIIOGKHAMHYSNRPKFKEDMLCNKCLNFRKLKCFRCGADKFDSEOVPPG
TTESQSVDDYCTIILRNIAPIHTVVDLSIMTASPYASLAVANNIRLIKDKQTQNRGF
AFVLSNDASQLLQILSLHPLKIDGKTIGVDFAKSARKDLVLSNDRVSFAFSVA
STATAAQSSTQSGGGSVDYSYLPQGGDGYAQTYSODYQOFOOAGGLES D
ASSAGTAVTTTSAAYVSPQSYNLTSPPGSPTEEAQFSTSTQAPAASTGYVP
GTYAVPDTSTQYDESSGYYPDTGLYDPNSQYYSNLTQOYLWDEKETVYPA
AESSHOOSGLPPAKGKKEKPKSKTAQOIAKDMERWAKSLKOKFNKFSQPVN
SLREERRESAADAAGFALFEKALAEROOLIPELVNRCDEENPLKGLVVAYSDS
DNEELVERLESEBEELADHWKMACLLCRQFENKDALVRHQSLDLHQNMDDIYRS
RLSEUELELELREKMYHRAAERREKIGIPEPPKRRKKQFDAGTVNMDYPTKDG
IDHSINGNKMLQAWREGSGLRKCGKITAPIEAQVRLKAGLGAAGKAYSGLSGADS
YKDAVRKAMFAFTEME"
566..589
misc_feature /gene="RBM5"
/notes="RNA-binding motif; unclassified site"
707..760
misc_feature /gene="RBM5"
/notes="zinc finger (C(x2)C(x10)C(x2)C) motif; DNA binding
site"
968..991
misc_feature /gene="RBM5"
/notes="RNA-binding motif; unclassified site"
2093..2164
misc_feature /gene="RBM5"
/notes="zinc finger (C(x2)C(x12)H(x6)H) motif; DNA binding
site"
2270..2320
misc_feature /gene="RBM5"
/notes="bipartite nuclear signal motif; signal-peptide
site"
BASE COUNT 900 a 705 c 822 g 708 t
ORIGIN

Query Match 19.3%; Score 518; DB 40; Length 3135;
Best Local Similarity 100.0%; Pred. No. 3.5e-276;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTTGGAGCTTCTGGGGCCAGAACCGCTACTGCTGCTTCGGTCTCTCTCTGGGAAAAAT 60
Db 43 GTTGGAGCTTCTGGGGCCAGAACCGCTACTGCTGCTCTCTCTCTGGGAAAAAT 102
Qy 61 AAAATTTGAACCTTTTGGAGCTGTGCTAAATCTTCAGTGGGACAATGGGTTTCAGACAA 120
Db 103 AAAATTTGAACCTTTTGGAGCTGTGCTAAATCTTCAGTGGGACAATGGGTTTCAGACAA 162
Qy 121 RAGAGTGAAGTGAACAGACGCGTAGTGGAGATACGGTTCCATCATAGACAGGATGACCG 180
Db 163 RAGAGTGAAGTGAACAGACGCGTAGTGGAGATACGGTTCCATCATAGACAGGATGACCG 222
Qy 181 TGATGAGCTGAATCCCGAAGCAGCGGAGGACTCAGATTACAAAAGATCTAGTGATGA 240
Db 223 TGATGAGCTGAATCCCGAAGCAGCGGAGGACTCAGATTACAAAAGATCTAGTGATGA 282
Qy 241 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATACAGATCCAGAGAGAGAGCGCTGA 300
Db 283 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATGACAGTCCAGAGAGAGAGCGCTGA 342
Qy 301 RAGAGGGAACAGTACCGATCCGAGATGCGTACCATTACAGATGGTACTATGGTGAGCA 360
Db 343 RAGAGGGAACAGTACCGATCCGAGATGCGTACCATTACAGATGGTACTATGGTGAGCA 402
Qy 361 CGACTATAGGCATGACATCAGTGAAGGAGGAGGAGCAAGACCATCATGCTGCGCGGCCT 420
Db 403 CGACTATAGGCATGACATCAGTGAAGGAGGAGGAGCAAGACCATCATGCTGCGCGGCCT 462
Qy 421 TCCCATCAACATCACAGAGACGATATTCGAGAAATGATGAGTCTCTTCGAAGGCCCTCA 480
Db 463 TCCCATCAACATCACAGAGACGATATTCGAGAAATGATGAGTCTCTTCGAAGGCCCTCA 522
Qy 481 GCCTCGGCTGTGAGGCTGATGAAGAGAGAAACAGGTG 518
```


Ddb	523	GCCTCGGATGTGGCGTGTGAAGAGAAACAGGTG	560
RESULT	3		
LOCUS	AF103802	3094 bp	mRNA
DEFINITION	Homo sapiens clone H37 unknown mRNA.	PRI	16-OCT-1999
ACCESSION	AF103802		
VERSION	AF103802.1	GI:5048971	
KEYWORDS	FLI_CDNA,		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	Oh,J.J., Grosshans D.R., Wong,S.G. and Slamon,D.J.		
JOURNAL	Identification of differentially expressed genes associated with		
MEDLINE	HER-2/neu overexpression in human breast cancer cells		
REFERENCE	Nucleic Acids Res. 27 (20), 4008-4017 (1999)		
AUTHORS	99428654		
TITLE	(bases 1 to 3094)		
JOURNAL	Oh,J.J., Wong,S.G. and Slamon,D.J.		
MEDLINE	Submitted (02-NOV-1998) Hematology/Oncology, UCLA School of		
REFERENCE	Medicine, 5-535 MRL, 675 Circle Drive South, Los Angeles, CA 90095,		
AUTHORS	USA		
JOURNAL	Location/Qualifiers		
MEDLINE	1. .3094		
REFERENCE	/organism="Homo sapiens"		
AUTHORS	/db_xref="taxon:9606"		
TITLE	/clone="H37"		
JOURNAL	/cell_line="MCF-7/HER2"		
MEDLINE	/cell_type="breast adenocarcinoma"		
REFERENCE	/note="MCF-7 cells retrovirally transformed with human		
AUTHORS	HER-2/neu full-length cDNA"		
TITLE	127. .2574		
JOURNAL	/note="expression increases in HER-2/neu overexpressing		
MEDLINE	cells; differentially expressed."		
REFERENCE	/codon_start=1		
AUTHORS	/product="unknown"		
TITLE	/protein_id="AAF02422.1"		
JOURNAL	/db_xref="GI:6048972"		
MEDLINE	/translation="MGSDKRVSRTERSGRYGSIIIDRRDRERSRSDYKRSSD		
REFERENCE	RGLPDTTYDYNDYDRERRNRDSDEGYHSDGYGBHDYRHDIHSERKSTIML		
AUTHORS	DRGLPTITTESDIRMEMSFEGPADVLNMRKTGVSRGFVFHELODATSWMEAN		
TITLE	KTEVIQGHKIAMYSNPFRFEDMLCNLFNRKRCFCRGADKFDSOEVPFG		
JOURNAL	TTESVQSYDYICDTHILRNIAPIHTVDISMTLPVSASLVANNIRIKDKTOQRGF		
MEDLINE	AFVLSAAMSQSOLLIQLSLPLPKIDGTIGVDFAKSKDLVLSDGNRVGSVA		
REFERENCE	STATAAWMSQTSQSGSVSYLQPGDGVAOYQSOYQFOQAAGLES		
AUTHORS	ASSAGTAVTTSTAAYVQSOPLYNOTSNPPGSTPEAOPTSTQPAAPTGVVP		
TITLE	GTKAVPDSTHYQDESQSYDDTLGLYDPNSQYNNLTQYLWGEKRTYVPA		
JOURNAL	AESSHQGUUPPAKEGKRPKTAQIAKDMERWAKSLNKONENKNGFPVN		
MEDLINE	SLREERESAADAAGALFKKGALAEQOLIPELVNNGDENLPRKLGLVAAYS		
REFERENCE	DNEELVESELEKLMACILCRQFFPNKDALVRHLQSLDLHKNMIDYRS		
AUTHORS	LRSQEALRELREMKYRDRAAREKRYGIPPEPPKKQFDAGTVNYEQPTDG		
TITLE	IDSHNIGNKMQAWREGSGLGRKCQGITAPIEAQVRLKAGLGKAGSAYLSGADS		
JOURNAL	YKDAVRKAMEARFEME"		
MEDLINE	BASE COUNT	879 a	702 c
ORIGIN	811 g	701 t	1 others
Query Match	17.4%;	Score 467;	DB 40; Length 3094;
Best Local Similarity	99.8%;	Pred. No. 6.9e-248;	
Matches	517; Conservative	0; Mismatches	1; Indels
	0; Gaps	0;	
Qy	1	GTGGAGGTCCTGGGCCGACAGCCGCTACTCCTCTCGGTCTCTCTCGGAAAAAT	60
Dd	21	GTGGAGGTCCTGGGCCGACAGCCGCTACTCCTCTCGGTCTCTCTCGGAAAAAT	80
Qy	61	AAAATTGAACCTTTTGAGCTGTGTGCTAAATCTTCAGTGGGACAATGGTTTCAGACAA	120
b	81	AAAATTGAACCTTTTGAGCTGTGTGCTAAATCTTCAGTGGGACAATGGTTTCAGACAA	140

Db 60 ATTGATGGTCTTCGAGGTAGTGTCTTAAGTGAAGTGTCTTCTCAGTTGTCTCAAGATAATGTT 1

RESULT 7
A77777/c 132 bp DNA PAT 19-OCT-1999
DEFINITION Sequence 484 from Patent EP0587279.
ACCESSION A77777
VERSION A77777.1 GI:6089442
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 132)
Sibson,D.R. and Hadfield,K.M.
HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,
TITLE PLACENTA OR BONE MARROW AND THEIR USE
JOURNAL Patent: EP 0587279-A 16-MAR-1994;
MEDICAL RES COUNCIL (GB)
FEATURES Location/Qualifiers
source 1. .132
/organism="unidentified"
/db_xref="taxon:32644" 37 t

BASE COUNT 49 a 26 c 20 g 37 t
ORIGIN

Query Match 4.5%; Score 120; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.3e-55;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1622 AGTGGCATTACATCTCAATCTCTTCGAGAAATGCCCTTTATAGTCGTTCAAAGCAAGTC 1681
|||||
Db 120 AGTGGCATTACATCTCAATCTCTTCGAGAAATGCCCTTTATAGTCGTTCAAAGCAAGTC 61
|||||

Qy 1682 ATTGATGGTCTTCGAGGTAGTGTCTTAAGTGAAGTGTCTTCAGTTGTCAAGATAATGTT 1741
|||||
Db 60 ATTGATGGTCTTCGAGGTAGTGTCTTAAGTGAAGTGTCTTCAGTTGTCAAGATAATGTT 1

RESULT 8
SCU96099 1115 bp mRNA INV 15-APR-1999
LOCUS Sarcophaga crassipalpis 23kDa heat shock protein SCHSP23 mRNA,
DEFINITION complete cds.
ACCESSION U96099
VERSION U96099.2 GI:4584801
KEYWORDS Sarcophaga crassipalpis.
SOURCE Sarcophaga crassipalpis.
ORGANISM Sarcophaga crassipalpis.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Oestroidea; Sarcophagidae; Sarcophaga.
REFERENCE 1 (bases 1 to 823)
AUTHORS Flanagan,R.D., Tammariello,S.P., Joplin,K.H., Cikra-Ireland,R.A.,
Yocum,G.D. and Denlinger,D.L.
TITLE Diapause-specific gene expression in pupae of the flesh fly
Sarcophaga crassipalpis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (10), 5616-5620 (1998)
MEDLINE 98245129
REFERENCE 2 (bases 1 to 823)
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.
TITLE Upregulation of a 23 kDa small heat shock protein transcript during
pupal diapause in the flesh fly, Sarcophaga, crassipalpis
JOURNAL Insect Biochem. Mol. Biol. 28 (9), 677-682 (1998)
MEDLINE 98428060
REFERENCE 3 (bases 1 to 823)
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1997) Entomology, Ohio State University, 1735
MEDLINE Neil Ave, Columbus, OH 43210-1220, USA
REFERENCE 4 (bases 1 to 1115)
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.

Direct Submission
Submitted (15-APR-1999) Entomology, Ohio State University, 1735
JOURNAL Neil Ave, Columbus, OH 43210-1220, USA
REMARK Sequence, update by submitter
COMMENT On Apr 15, 1999 this sequence version replaced gi:2058736.
FEATURES
source 1. .1115
Location/Qualifiers
/organism="Sarcophaga crassipalpis"
/db_xref="taxon:59312"
88. .705
/codon_start=1
/product="23kDa heat shock protein SCHSP23"
/protein_id="AAC63387.1"
/translation="MAGLPLLLSLIDELRESNPSYYGNDGFLGSLPYLTHRPQREP
AHNLVGYSLPLSLASRLNEHQVARRGEGKEKRGVSGKDFQVCMQVADQFKPSLNV
KVVDNCIVIEGKHEEREDOHGFIOHFVRVLPKGYDADKVVSTLSSDGLTVSVPK
POEEKSNERVIOQQTGPAHLNVKONTQEKVGKGANGEPQEEKK"
BASE COUNT 430 a 185 c 192 g 308 t
ORIGIN

Query Match 1.7%; Score 47; DB 35; Length 1115;
Best Local Similarity 100.0%; Pred. No. 7.2e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2643 ACCCCCAA 2689
|||||
Db 1027 ACCCCCAA 1073
|||||

RESULT 9
RATL1NAT 2305 bp mRNA ROD 27-APR-1993
LOCUS Rat L-type neutral amino acid transporter mRNA, complete cds.
DEFINITION M77345
ACCESSION M77345.1 GI:205238
VERSION L-type neutral amino acid transporter.
KEYWORDS Rattus norvegicus (strain Wistar) cDNA to mRNA.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 2305)
AUTHORS Tate,S.S., Yan,N. and Udenfriend,S.
TITLE Expression cloning of a Na⁺-independent neutral amino acid
transporter from rat kidney
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89, 1-5 (1992)
MEDLINE 92107996
FEATURES Location/Qualifiers
source 1. .2305
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
39. .2090
/codon_start=1
/product="L-type neutral amino acid transporter"
/protein_id="AA41544.1"
/translation="MNEDKDKRDSIQMSKMGCRNNGFVNEDIQEODPDSDRTPQSN
AVISIPAPEPQLKVVRYPYAGMPKEVLFQSGQARYRVPREILFWLTVVSVELLIGAT
AIIISPKCLDWQAGPMYQIYVPSFKDSKDGNDLKGIOEKLDYITALNKTITWIT
WFQSPKLSRGYVDEFEIDIFGTMTDFENLVAHVHDKGLKLIIDFIPNHTSDKHP
SFQSSRTSGKYTDYIWNHCTHANGVTTPNNMLSVYGVNSWQFDEERKOCYFQFL
KEQPLNFRNPVQEIKELIKFWLSKGVDFSDAVKFLLEAKDLNEIQVNTSQIP
DIVTFSELDHFTTQVGMHDLVDFQTMNQFSRPGRFPMGTVEVSASERTMY
YYGLSFQDAEPFNKILATDLSGHTVYEAITSWMENMPGKWNMIGGPTSRU
TSRVGEYVYNNMMLLFTPGTPIYTGEEIGMGDISITNLNERYDINALLSKSPGW
DNSSAGFTTEANHTLPTNSDYHTVNDVQTPQSSALRYQDLSLHARELLSRGW
FCLLRDNHVVYTRLEDGIDKVLVNFGEISTVLNQETISDPTKIRLSTNP
C"
BASE COUNT 691 a 540 c 532 g 542 t
ORIGIN

```
Query Match          1.7%: Score 47: DB 12: Length 2305;
Best Local Similarity 100.0%: Pred. No. 7.2e-15;
Matches 47: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2643 ACCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||||
DB 2254 ACCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2300

RESULT 10
AC013452      184553 bp      DNA      HTG      12-NOV-1999
LOCUS      Homo sapiens chromosome 15 clone RP11-325E5 map 15q21.1, LOW-PASS
DEFINITION      SEQUENCE SAMPLING.
AC013452
ACCESSION      AC013452.1 GI:6382485
VERSION
KEYWORDS
SOURCE      HTG; HTGS_PHASE0.
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184553)
Roven,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117region
Unpublished
2 (bases 1 to 184553)
Roven,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (12-NOV-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@u.washington.edu
----- Summary Statistics
Sequencing vector: pUC18: L08752
Chemistry: Big Dye terminators
Assembly program: Phrap; version 0.990399.
* NOTE: this record contains 192 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 914: contig of 914 bp in length
* gap of unknown length
* 915 1802: contig of 888 bp in length
* gap of unknown length
* 1803 2360: contig of 558 bp in length
* gap of unknown length
* 2361 3248: contig of 888 bp in length
* gap of unknown length
* 3249 3769: contig of 521 bp in length
* gap of unknown length
* 3770 4671: contig of 902 bp in length
* gap of unknown length
* 4672 5662: contig of 991 bp in length
* gap of unknown length
* 5663 6561: contig of 899 bp in length
* gap of unknown length

7466: contig of 905 bp in length
8367: contig of 901 bp in length
9293: contig of 926 bp in length
10195: contig of 902 bp in length
10662: contig of 467 bp in length
11812: contig of 1150 bp in length
12734: contig of 922 bp in length
13630: contig of 896 bp in length
14561: contig of 931 bp in length
15459: contig of 898 bp in length
16461: contig of 1002 bp in length
17358: contig of 897 bp in length
18322: contig of 964 bp in length
19222: contig of 900 bp in length
20193: contig of 971 bp in length
21221: contig of 1028 bp in length
21727: contig of 506 bp in length
22830: contig of 1103 bp in length
23796: contig of 966 bp in length
24970: contig of 1174 bp in length
25769: contig of 799 bp in length
26743: contig of 974 bp in length
27692: contig of 949 bp in length
28682: contig of 990 bp in length
29675: contig of 993 bp in length
30575: contig of 900 bp in length
31555: contig of 980 bp in length
32453: contig of 898 bp in length
33425: contig of 972 bp in length
34346: contig of 921 bp in length
35510: contig of 1164 bp in length
36420: contig of 910 bp in length
36826: contig of 406 bp in length
38207: contig of 1381 bp in length
39027: contig of 820 bp in length
39931: contig of 904 bp in length
gap of unknown length
```

* 39932 40863: contig of 932 bp in length
* gap of unknown length
* 40864 41894: contig of 1031 bp in length
* gap of unknown length
* 41895 42767: contig of 873 bp in length
* gap of unknown length
* 42768 43667: contig of 900 bp in length
* gap of unknown length
* 43668 44870: contig of 1203 bp in length
* gap of unknown length
* 44871 45843: contig of 973 bp in length
* gap of unknown length
* 45844 46783: contig of 940 bp in length
* gap of unknown length
* 46784 47851: contig of 1068 bp in length
* gap of unknown length
* 47852 48830: contig of 979 bp in length
* gap of unknown length
* 48831 50308: contig of 1478 bp in length
* gap of unknown length
* 50309 51547: contig of 1239 bp in length
* gap of unknown length
* 51548 52617: contig of 1070 bp in length
* gap of unknown length
* 52618 53551: contig of 934 bp in length
* gap of unknown length
* 53552 54449: contig of 898 bp in length
* gap of unknown length
* 54450 55434: contig of 985 bp in length
* gap of unknown length
* 55435 56625: contig of 1191 bp in length
* gap of unknown length
* 56626 57543: contig of 918 bp in length
* gap of unknown length
* 57544 58518: contig of 975 bp in length
* gap of unknown length
* 58519 59476: contig of 958 bp in length
* gap of unknown length
* 59477 60932: contig of 1456 bp in length
* gap of unknown length
* 60933 61859: contig of 927 bp in length
* gap of unknown length
* 61860 62750: contig of 891 bp in length
* gap of unknown length
* 62751 63683: contig of 933 bp in length
* gap of unknown length
* 63684 64581: contig of 898 bp in length
* gap of unknown length
* 64582 65562: contig of 981 bp in length
* gap of unknown length
* 65563 66461: contig of 899 bp in length
* gap of unknown length
* 66462 67594: contig of 1133 bp in length
* gap of unknown length
* 67595 69414: contig of 1820 bp in length
* gap of unknown length
* 69415 70498: contig of 1084 bp in length
* gap of unknown length
* 70499 71852: contig of 1354 bp in length
* gap of unknown length
* 71853 72767: contig of 915 bp in length
* gap of unknown length
* 72768 73669: contig of 902 bp in length
* gap of unknown length
* 73670 74584: contig of 915 bp in length
* gap of unknown length
* 74585 75486: contig of 902 bp in length
* gap of unknown length

Query Match 1.7%; Score 47; DB 43; Length 184553;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2643 ACCCCCAA 2689

Db 115012 ACCCCCAA 115058

RESULT 11
A69673
LOCUS A69673 1305 bp DNA PAT 07-MAY-1999
DEFINITION Sequence 3 from Patent WO9807872.
ACCESSION A69673
VERSION A69673.1 GI:4774286
KEYWORDS
SOURCE Trichomonas vaginalis.
ORGANISM Trichomonas vaginalis.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Campbell,R.S., Coombs,G.H., Mottram,J.C. and Pritchard,D.J.
TITLE HOMOCYSTEINE DESULPHURASE FROM THE PROTOZOAN TRICHOMONAS VAGINALIS
JOURNAL Patent: WO 9807872-A 26-FEB-1998;
CAMPBELL ROBERT STEWART (GB)
FEATURES
source
1. .1305
/organism="Trichomonas vaginalis"
/db_xref="taxon:5722"
/cell_line="G3"
/clone="MGL2"
16. .1212
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB42512.1"
/db_xref="GI:4774287"
/translation="MSGHAIDPTHTDLSIHANPKDQFGAIVAPIYQTSFLFNCD
OGGARFGGKEAGYMTIRGNPTNSALEGKIAKLEHAEACAATAAGMGALAAASVWTF
AGDHLISDDCLYGCTHALFEHLKFGVEVDMAVPGNIEKHLKPNTRIVVETPA
NPTLVINIDIAVKQASKOKDLIVIVNTFASPLNPDLGLVDIVVHSATKVINHT
DVAVGLVCSRAIIIAKVSQGIKIDITGAIISPDAHLITRGTLITLDMRVKRAENAK
VAEFUHEKAVKVVYPLDPHPGHEIAKKMKFMFGSMIAFDVDDGLEKAKKVLDDNC
VSLAYSLGGPESLIQHPASMTGHVGPKEEREAAGLTNDNLRLSVGCGENVQDIIDDLKQ
ALDLVL"

CDS
1
16. .1212
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB42512.1"
/db_xref="GI:4774287"
/translation="MSGHAIDPTHTDLSIHANPKDQFGAIVAPIYQTSFLFNCD
OGGARFGGKEAGYMTIRGNPTNSALEGKIAKLEHAEACAATAAGMGALAAASVWTF
AGDHLISDDCLYGCTHALFEHLKFGVEVDMAVPGNIEKHLKPNTRIVVETPA
NPTLVINIDIAVKQASKOKDLIVIVNTFASPLNPDLGLVDIVVHSATKVINHT
DVAVGLVCSRAIIIAKVSQGIKIDITGAIISPDAHLITRGTLITLDMRVKRAENAK
VAEFUHEKAVKVVYPLDPHPGHEIAKKMKFMFGSMIAFDVDDGLEKAKKVLDDNC
VSLAYSLGGPESLIQHPASMTGHVGPKEEREAAGLTNDNLRLSVGCGENVQDIIDDLKQ
ALDLVL"

BASE COUNT 374 a 376 c 269 g 286 t
ORIGIN
Query Match 1.7%; Score 46; DB 5; Length 1305;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2644 CCCCCAAA 2689
Db 1257 CCCCCAAA 1302

RESULT 12
A69681
LOCUS A69681 1306 bp DNA PAT 07-MAY-1999
DEFINITION Sequence 11 from Patent WO9807872.
ACCESSION A69681
VERSION A69681.1 GI:4774294
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1306)
AUTHORS Campbell,R.S., Coombs,G.H., Mottram,J.C. and Pritchard,D.J.
TITLE HOMOCYSTEINE DESULPHURASE FROM THE PROTOZOAN TRICHOMONAS VAGINALIS
JOURNAL Patent: WO 9807872-A 26-FEB-1998;
CAMPBELL ROBERT STEWART (GB)
FEATURES
source
1. .1306
/organism="unidentified"
/db_xref="taxon:32644"
17. .1213
/note="unnamed protein product"
/codon_start=1

Page 11

LYIDKEPTVNNNNNNSNNSNNSNNSNNSNNSNNDQIDLVIYSDDKKSSLIY

PNYASNFKGRFSYQTYEIVNSANLMAIKDLDYGDQLGPLNLAKVSMQIDENKILVGL
CASNKLI RTNNSSKVYNEIFNGIINELSONHGVALDVFSNPNGTSPNLTLQNNNNN
NNNSGGSNSSKKDLEGRVEIINQADLTTKLRGVLIGSVEAFRRNKVADSIIKIGGMPIS

NIELFLDLIGNCSTTITSRAPOEGTVANWNACKYIMNWDIWRLTTPALQRHVI
NGYNSLIVNNIQRFRNIDSLRKVNVIQEIFILSSSTNEEPLPESVASSVINVLNLI
SYGGLIEDDIROIISAFILSHLHKDIPTPSSSSSSSSSTSTSSRRKSIHRSKLATMELS

SSSGAGGAGTIVPEHLLIRITPEPSKLCYERAAQLILKRSYEDNVXYLEQKQO
LQNGAGDDELLSLNVLNFINNNKNNNNNNPLSTPQTISRVSYSNLSNLS
SSSSTLVNSNNNNNNPTSGASITITKFNWKEFKTEFAVTTGADENSTG
ATDAQAAKLLKKRMSIOSPOFKNLGGDDSVTTPGSLHARVGTMMDDSKLN
GALGGGGGGGGYGVGNQPINFLNDELPELKSQFAPEESQYTMLTTFYI

CQIMLSAMRKTSAISIEMVLEGAPTATDEEFIYHSRILLDLMYVETNITKTEF

TVQSLYKSLNRILYLINHTTDTLSFVANHIINHQRIIFSENNLDSDFMNAFCYPLY
KLVIDSQHEHVDNSIKLWRLLSLKTSSYIESLATVLQKLVSSGNSQROSEIIDLKPG
FELLRNTSGNGAFNDEFLWINDNIQTITQVFEENPKKQHLSPFKNNEKKKHSSEHTLP

FNAIQWENMRAQITRERAVWGPSEPHPLDKWKLDSTEGPYRMRKKMEKNYNFYKNYPY

SSITNNNNNNNNNNNNNNNTITKSTQNANNNNNANNQSTSSSSNTTTTTT
 PQOSSQIKVSSPELSSNEITPTSPVQSSSEDFKSKLQSTVEGQLSRNPSSSEL
 FNDNSSTISEENSSLTSASTLSPPPPSTQTTTTTTTSTPTQSSVATTTTGTNTNEVD

ILIFCPVYMYIFDGYKDENTGDISVEEIKINSEWLPECTVLPMKKIIHYFLKWAYE

QGITGGQTGNDNDHHGGGGGGRGVRDRFTSIWRKSPLTLKWQQGQISNFQYLMHLNT
LGRSYNDLTQYVPFPVWLSDYEESEELDIDDPKVYRDLSKPMGALEESRAQKFRFERFE

SSITEAWASSSQSTGVVMELIPEFYLDDEFVLVNNKFNFGTKQGGEPIDDIILPPWA

IDAISDPVEKAATAIAQINNFGQTPKQLFDKPPHPRKNATLMGLPFYAKALTGNFIKDIDG
EPVQGIRLINDRATCVGFKNVLLPPNHSKYMLWGLPDGSI RYNTGDKIKVLEDHHDGPP

SDDRTCIWDLNRLCYVRS LDAHEGPISCIGIHD TTGEIVVCSGTTISVYTVNGELLI
NYKTSQIANDQITCCIWSKGPWLGENVLLTGRHDGKVVKWGLETRLL.PDNNNSNNNNN

DITGRVCMWSNEASQVKQRGWINTDIRGILQSEKK "
12469. .12472

/note="DpnII site at which plasmid pRHI30 was inserted in the REMI mutant 60TC5"

```
/gene="lvsa"
12629
```

a	1757 c	1815 g	4520 t
---	--------	--------	--------

	1.78;	Score	46;	DB	35;	Length	12971;
100	08.	2003	NIC	2	50-14.		

```

servative 0; Mismatches 0; Indels 0; Gaps 0;

```

AAAAAAAAAAAAAAAAAAAAAAAAAAAA 445

28852 bp DNA HTG 15-SEP-1999
 iens chromosome 5 clone CIT-HSPC_349C3, *** SEQUENCING IN
 ***, 32 unordered pieces.
 .1 GI:5882694
 S_PHASE1.

[illegible]

```

12409..12472
/gene="lvsA"
/feature="DpnII site at which plasmid pRH30 was inserted in
the REMI mutant 60TC5"
12604..12971
/gene="lvsA"
12629
polyA_site
BASE COUNT 4879 a 1757 c 1815 g 4520 t
ORIGIN
Query Match 1.7% Score 46; DB 35; Length 12971;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Db 400 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 445
|||||
RESULT 16
AC010238 AC010238 28862 bp DNA HTG 15-SEP-1999
LOCUS Homo sapiens Chromosome 5 Clone CIT-HSPC_349C3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 32 unordered pieces.
AC010238 AC010238
AC010238 AC010238.1 GI:5882694
KEYWORDS HTG; HTGS_PHASE1.

```

[illegible]

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 28862)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 28862)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT www.jgi.doe.gov.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 725: contig of 725 bp in length
 * gap of unknown length
 * 726 1942: contig of 1217 bp in length
 * gap of unknown length
 * 1943 2512: contig of 570 bp in length
 * gap of unknown length
 * 2513 5735: contig of 3223 bp in length
 * gap of unknown length
 * 5736 6514: contig of 779 bp in length
 * gap of unknown length
 * 6515 7081: contig of 567 bp in length
 * gap of unknown length
 * 7082 7787: contig of 706 bp in length
 * gap of unknown length
 * 7788 8551: contig of 764 bp in length
 * gap of unknown length
 * 8552 9198: contig of 647 bp in length
 * gap of unknown length
 * 9199 9983: contig of 785 bp in length
 * gap of unknown length
 * 9984 10428: contig of 445 bp in length
 * gap of unknown length
 * 10429 10686: contig of 258 bp in length
 * gap of unknown length
 * 10687 11548: contig of 862 bp in length
 * gap of unknown length
 * 11549 12265: contig of 717 bp in length
 * gap of unknown length
 * 12266 13168: contig of 903 bp in length
 * gap of unknown length
 * 13169 13854: contig of 686 bp in length
 * gap of unknown length
 * 13855 14636: contig of 782 bp in length
 * gap of unknown length
 * 14637 15872: contig of 1236 bp in length
 * gap of unknown length
 * 15873 16842: contig of 970 bp in length
 * gap of unknown length
 * 16843 17579: contig of 737 bp in length
 * gap of unknown length
 * 17580 18803: contig of 1224 bp in length
 * gap of unknown length
 * 18804 19386: contig of 583 bp in length
 * gap of unknown length
 * 19387 20433: contig of 1047 bp in length
 * gap of unknown length
 * 20434 21170: contig of 737 bp in length
 * gap of unknown length
 * 21171 21263: contig of 93 bp in length
 * gap of unknown length
 * 21264 21913: contig of 650 bp in length

* 21914 23120: contig of 1207 bp in length
 * gap of unknown length
 * 23121 24749: contig of 1629 bp in length
 * gap of unknown length
 * 24750 25568: contig of 819 bp in length
 * gap of unknown length
 * 25569 26585: contig of 1017 bp in length
 * gap of unknown length
 * 26586 27627: contig of 1042 bp in length
 * gap of unknown length
 * 27628 28862: contig of 1235 bp in length.
 FEATURES Location/Qualifiers
 source 1..28862
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CIT-RSPC_349C3"
 BASE COUNT 7289 a 7328 c 7041 g 6799 t 405 others
 ORIGIN
 QY 2644 CCCCACAAAAA
 DB 1420 CCCCACAAAAA
 Query Match 1.7%; Score 46; DB 41; Length 28862;
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2644 CCCCACAAAAA
 DB 1420 CCCCACAAAAA
 RESULT 17
 AC016250/c
 LOCUS AC016250 78250 bp DNA HTG 23-NOV-1999
 DEFINITION Homo sapiens chromosome 15 clone RP11-516P9 map 15, LOW-PASS
 SEQUENCE SAMPLING.
 ACCESSION AC016250
 VERSION AC016250.1 GI:6466544
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 78250)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 78250)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckhgalter, B.,
 Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A.,
 Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehotzky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 TITLE Direct Submission
 JOURNAL Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submission@genome.wi.mit.edu


```
----- Project Information
Center project name: L4978
Center clone name: 516_P_9
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 981: contig of 981 bp in length
* gap of unknown length
* 982 1969: contig of 988 bp in length
* gap of unknown length
* 1970 2951: contig of 982 bp in length
* gap of unknown length
* 2952 3967: contig of 1016 bp in length
* gap of unknown length
* 3968 4876: contig of 909 bp in length
* gap of unknown length
* 4877 5699: contig of 823 bp in length
* gap of unknown length
* 5700 6669: contig of 970 bp in length
* gap of unknown length
* 6670 7627: contig of 958 bp in length
* gap of unknown length
* 7628 8552: contig of 925 bp in length
* gap of unknown length
* 8553 9533: contig of 981 bp in length
* gap of unknown length
* 9534 10496: contig of 963 bp in length
* gap of unknown length
* 10497 11483: contig of 987 bp in length
* gap of unknown length
* 11484 12457: contig of 974 bp in length
* gap of unknown length
* 12458 13440: contig of 983 bp in length
* gap of unknown length
* 13441 14419: contig of 979 bp in length
* gap of unknown length
* 14420 15392: contig of 973 bp in length
* gap of unknown length
* 15393 16338: contig of 946 bp in length
* gap of unknown length
* 16339 17295: contig of 957 bp in length
* gap of unknown length
* 17296 18288: contig of 993 bp in length
* gap of unknown length
* 18289 19227: contig of 939 bp in length
* gap of unknown length
* 19228 20137: contig of 910 bp in length
* gap of unknown length
* 20138 21111: contig of 974 bp in length
* gap of unknown length
* 21112 22088: contig of 977 bp in length
* gap of unknown length
* 22089 23089: contig of 1001 bp in length
* gap of unknown length
* 23090 24039: contig of 950 bp in length
* gap of unknown length
* 24040 24951: contig of 912 bp in length
* gap of unknown length
* 24952 25924: contig of 973 bp in length
* gap of unknown length
* 25925 26930: contig of 1006 bp in length
* gap of unknown length
* 26931 27913: contig of 983 bp in length
* gap of unknown length
*
* 27914 28882: contig of 969 bp in length
* gap of unknown length
* 28883 29825: contig of 943 bp in length
* gap of unknown length
* 29826 30825: contig of 1000 bp in length
* gap of unknown length
* 30826 31777: contig of 952 bp in length
* gap of unknown length
* 31778 32754: contig of 977 bp in length
* gap of unknown length
* 32755 33733: contig of 979 bp in length
* gap of unknown length
* 33734 34725: contig of 992 bp in length
* gap of unknown length
* 34726 35667: contig of 942 bp in length
* gap of unknown length
* 35668 36638: contig of 971 bp in length
* gap of unknown length
* 36639 37590: contig of 952 bp in length
* gap of unknown length
* 37591 38548: contig of 938 bp in length
* gap of unknown length
* 38549 39502: contig of 954 bp in length
* gap of unknown length
* 39503 40448: contig of 946 bp in length
* gap of unknown length
* 40449 41431: contig of 983 bp in length
* gap of unknown length
* 41432 42427: contig of 996 bp in length
* gap of unknown length
* 42428 43424: contig of 997 bp in length
* gap of unknown length
* 43425 44355: contig of 931 bp in length
* gap of unknown length
* 44356 45315: contig of 960 bp in length
* gap of unknown length
* 45316 46264: contig of 949 bp in length
* gap of unknown length
* 46265 47222: contig of 958 bp in length
* gap of unknown length
* 47223 48185: contig of 963 bp in length
* gap of unknown length
* 48186 49143: contig of 958 bp in length
* gap of unknown length
* 49144 50119: contig of 976 bp in length
* gap of unknown length
* 50120 51091: contig of 972 bp in length
* gap of unknown length
* 51092 52032: contig of 941 bp in length
* gap of unknown length
* 52033 52985: contig of 953 bp in length
* gap of unknown length
* 52986 53943: contig of 958 bp in length
* gap of unknown length
* 53944 54947: contig of 1004 bp in length
* gap of unknown length
* 54948 55933: contig of 986 bp in length
* gap of unknown length
* 55934 56879: contig of 946 bp in length
* gap of unknown length
* 56880 57886: contig of 1007 bp in length
* gap of unknown length
* 57887 58879: contig of 993 bp in length
* gap of unknown length
* 58880 59864: contig of 985 bp in length
* gap of unknown length
* 59865 60834: contig of 970 bp in length
* gap of unknown length
* 60835 61817: contig of 983 bp in length
* gap of unknown length
* 61818 62801: contig of 984 bp in length
* gap of unknown length
* 62802 63787: contig of 986 bp in length
* gap of unknown length
```

* 63788	64726:	gap of unknown length	4117	4612:	contig of 496 bp in length
* *		contig of 939 bp in length	4613	4622:	gap of unknown length
* *		gap of unknown length	4623	5136:	contig of 514 bp in length
* *	64727	65673: contig of 947 bp in length	5137	5146:	gap of unknown length
* *		gap of unknown length	5147	5661:	contig of 515 bp in length
* *	65674	66654: contig of 981 bp in length	5662	5671:	gap of unknown length
* *		gap of unknown length	5672	6185:	contig of 514 bp in length
* *	66655	67598: contig of 944 bp in length	6186	6195:	gap of unknown length
* *		gap of unknown length	6196	6691:	contig of 496 bp in length
* *	67599	68566: contig of 968 bp in length	6692	6701:	gap of unknown length
* *		gap of unknown length	6702	7218:	contig of 517 bp in length
* *	68567	69538: contig of 972 bp in length	7219	7228:	gap of unknown length
* *		gap of unknown length	7229	7741:	contig of 513 bp in length
* *	69539	70523: contig of 985 bp in length	7742	7751:	gap of unknown length
* *			7752	8264:	contig of 513 bp in length
* *		1.7%; Score 45; DB 44; Length 78250;	8265	8274:	gap of unknown length
* *	Query Match	100.0%; Pred. No. 2.5e-14;	8275	8793:	contig of 519 bp in length
* *	Best Local Similarity	0; Mismatches 0; Indels 0; Gaps 0;	8794	8803:	gap of unknown length
* *	Matches 46; Conservative		8804	9334:	contig of 531 bp in length
Qy 2644	CCCCCAA	2689	9345	9344:	gap of unknown length
			9345	9856:	contig of 512 bp in length
Db 36931	CCCCCAA	36866	9857	9866:	gap of unknown length
			9867	10382:	contig of 516 bp in length
			10383	10392:	gap of unknown length
RESULT 18			10393	11113:	contig of 721 bp in length
AC017106/c			11114	11123:	gap of unknown length
LOCUS	AC017106	107475 bp	11124	11638:	contig of 515 bp in length
DEFINITION	Homo sapiens clone RP11-565G4, LOW-PASS SEQUENCE SAMPLING.	HTG	11639	11648:	gap of unknown length
ACCESSION	AC017106		11649	12358:	contig of 710 bp in length
VERSION	AC017106.1	GI:6554034	12359	12368:	gap of unknown length
KEYWORDS	HTG: HTGS_PHASE0.		12369	12885:	contig of 517 bp in length
SOURCE	human.		12886	12895:	gap of unknown length
ORGANISM	Homo sapiens		12896	13410:	contig of 515 bp in length
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		13411	13420:	gap of unknown length
REFERENCE	1 (bases 1 to 107475)		13421	13944:	contig of 524 bp in length
AUTHORS	Waterston,R.H.		13945	13954:	gap of unknown length
TITLE	The sequence of Homo sapiens clone		13955	14473:	contig of 519 bp in length
JOURNAL	Unpublished		14474	14483:	gap of unknown length
REFERENCE	2 (bases 1 to 107475)		14484	15000:	contig of 517 bp in length
AUTHORS	Waterston,R.H.		15001	15010:	gap of unknown length
TITLE	Direct Submission		15011	15471:	contig of 461 bp in length
JOURNAL	Submitted (09-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		15472	15481:	gap of unknown length
COMMENT	Center project name: H_NH0567N19.		15482	15994:	contig of 513 bp in length
	* NOTE: This record contains 186 individual		15995	16004:	gap of unknown length
	* sequencing reads that have not been assembled into		16005	16520:	contig of 516 bp in length
	* contigs. Runs of N are used to separate the reads		16521	16530:	gap of unknown length
	* and the order in which they appear is completely		16531	17045:	contig of 515 bp in length
	* arbitrary. Low-pass sequence sampling is useful for		17046	17578:	contig of 523 bp in length
	* identifying clones that may be gene-rich and allows		17579	17588:	gap of unknown length
	* overlap relationships among clones to be deduced.		17589	18298:	contig of 710 bp in length
	* However, it should not be assumed that this clone		18299	18308:	gap of unknown length
	* will be sequenced to completion. In the event that		18309	18777:	contig of 469 bp in length
	* the record is updated, the accession number will		18778	18787:	gap of unknown length
	* be preserved.		18788	19519:	contig of 732 bp in length
			19520	19529:	gap of unknown length
			19530	20017:	contig of 488 bp in length
			20018	20027:	gap of unknown length
			20028	20324:	contig of 305 bp in length
			20329	20859:	contig of 517 bp in length
			20333	20869:	gap of unknown length
			20860	21398:	contig of 529 bp in length
			20870	21408:	gap of unknown length
			21399	21409:	contig of 538 bp in length
			21409	21946:	contig of 516 bp in length
			21947	21956:	gap of unknown length
			21957	22472:	contig of 516 bp in length
			22473	22482:	gap of unknown length
			22483	23008:	contig of 524 bp in length

* 23873 23882: gap of unknown length
* 23883 23997: contig of 515 bp in length
* 24398 24407: gap of unknown length
* 24924 24933: contig of 516 bp in length
* 24934 24941: gap of unknown length
* 25434 25441: contig of 508 bp in length
* 25442 25451: gap of unknown length
* 25452 25970: contig of 519 bp in length
* 25971 25980: gap of unknown length
* 25981 26498: contig of 518 bp in length
* 26499 26508: gap of unknown length
* 26509 27034: contig of 526 bp in length
* 27035 27044: gap of unknown length
* 27045 27558: contig of 514 bp in length
* 27559 27568: gap of unknown length
* 27569 28104: contig of 536 bp in length
* 28105 28114: gap of unknown length
* 28115 28624: contig of 510 bp in length
* 28625 28634: gap of unknown length
* 28635 29046: contig of 412 bp in length
* 29047 29056: gap of unknown length
* 29057 29772: contig of 716 bp in length
* 29773 29782: gap of unknown length
* 29783 30514: contig of 732 bp in length
* 30515 30524: gap of unknown length
* 30525 31038: contig of 514 bp in length
* 31039 31048: gap of unknown length
* 31049 31787: contig of 739 bp in length
* 31788 31797: gap of unknown length
* 31798 32300: contig of 503 bp in length
* 32301 32310: gap of unknown length
* 32311 32826: contig of 516 bp in length
* 32827 32836: gap of unknown length
* 32837 33355: contig of 519 bp in length
* 33356 33365: gap of unknown length
* 33366 33884: contig of 519 bp in length
* 33885 33894: gap of unknown length
* 33895 34300: contig of 406 bp in length
* 34301 34310: gap of unknown length
* 34311 34819: contig of 509 bp in length
* 34820 34829: gap of unknown length
* 34830 35346: contig of 517 bp in length
* 35347 35356: gap of unknown length
* 35357 35796: contig of 440 bp in length
* 35797 35806: gap of unknown length
* 35807 36324: contig of 518 bp in length
* 36325 36334: gap of unknown length
* 36335 36843: contig of 509 bp in length
* 36844 36853: gap of unknown length
* 36854 37237: contig of 444 bp in length
* 37238 37307: gap of unknown length
* 37308 37824: contig of 517 bp in length
* 37825 37834: gap of unknown length
* 37835 38347: contig of 513 bp in length
* 38348 38357: gap of unknown length
* 38358 38788: contig of 431 bp in length
* 38789 38798: gap of unknown length
* 38799 39312: contig of 514 bp in length
* 39313 39322: gap of unknown length
* 39323 39841: contig of 519 bp in length
* 39842 39851: gap of unknown length
* 39852 40213: contig of 362 bp in length
* 40214 40223: gap of unknown length
* 40224 40737: contig of 514 bp in length
* 40738 40747: gap of unknown length
* 40748 41263: contig of 516 bp in length
* 41264 41273: gap of unknown length
* 41274 41785: contig of 512 bp in length
* 41786 41795: gap of unknown length
* 41796 42311: contig of 516 bp in length
* 42312 42321: gap of unknown length
* 42322 42561: contig of 240 bp in length
* 42562 42571: gap of unknown length

* 42572 43312: contig of 741 bp in length
* 43313 43322: gap of unknown length
* 43323 43841: contig of 519 bp in length
* 43842 43851: gap of unknown length
* 43852 44370: contig of 519 bp in length
* 44371 44380: gap of unknown length
* 44381 44898: contig of 518 bp in length

Query Match 1.7%; Score 46; DB 45; Length 107475;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
 |||||
Db 106788 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106743

RESULT 19
AC010404
LOCUS AC010404 108064 bp DNA HTG 15-SEP-1999
DEFINITION Homo sapiens chromosome 5 clone CITB-H1_2149024, *** SEQUENCING IN
 PROGRESS ***, 87 unordered pieces.
ACCESSION AC010404
VERSION AC010404.1 GI:5882532
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108064)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108064)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 www.jgi.doe.gov.
COMMENT * NOTE: This is a 'working draft' sequence. It currently
 * consists of 87 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 3167: contig of 3167 bp in length
 * 3168 5734: gap of unknown length
 * 5735 6511: contig of 2567 bp in length
 * 6512 10376: gap of unknown length
 * 10377 14385: contig of 777 bp in length
 * 14386 17326: contig of 3865 bp in length
 * 17327 22463: gap of unknown length
 * 22464 28511: contig of 4009 bp in length
 * 28512 29102: gap of unknown length
 * 29103 36141: contig of 2941 bp in length
 * 36142 42474: gap of unknown length
 * 42475 43215: contig of 5137 bp in length
 * 43216 43853: contig of 6048 bp in length
 * gap of unknown length
 * contig of 591 bp in length
 * contig of 7039 bp in length
 * contig of 6333 bp in length
 * contig of 741 bp in length
 * contig of 638 bp in length
 * gap of unknown length

*	43854	44494:	contig of 641 bp in length	*
*		gap of unknown length		*
*	44495	45122:	contig of 628 bp in length	*
*		gap of unknown length		*
*	45123	45775:	contig of 653 bp in length	*
*		gap of unknown length		*
*	45776	46407:	contig of 632 bp in length	*
*		gap of unknown length		*
*	46408	47096:	contig of 689 bp in length	*
*		gap of unknown length		*
*	47097	47746:	contig of 650 bp in length	*
*		gap of unknown length		*
*	47747	48434:	contig of 688 bp in length	*
*		gap of unknown length		*
*	48435	49043:	contig of 609 bp in length	*
*		gap of unknown length		*
*	49044	49487:	contig of 444 bp in length	*
*		gap of unknown length		*
*	49488	50164:	contig of 677 bp in length	*
*		gap of unknown length		*
*	50165	51352:	contig of 1188 bp in length	*
*		gap of unknown length		*
*	51353	52008:	contig of 656 bp in length	*
*		gap of unknown length		*
*	52009	53126:	contig of 1118 bp in length	*
*		gap of unknown length		*
*	53127	53880:	contig of 754 bp in length	*
*		gap of unknown length		*
*	53881	54583:	contig of 703 bp in length	*
*		gap of unknown length		*
*	54584	55426:	contig of 843 bp in length	*
*		gap of unknown length		*
*	55427	55899:	contig of 473 bp in length	*
*		gap of unknown length		*
*	55900	56462:	contig of 563 bp in length	*
*		gap of unknown length		*
*	56463	57089:	contig of 627 bp in length	*
*		gap of unknown length		*
*	57090	57665:	contig of 576 bp in length	*
*		gap of unknown length		*
*	57666	58483:	contig of 818 bp in length	*
*		gap of unknown length		*
*	58484	59708:	contig of 1223 bp in length	*
*		gap of unknown length		*
*	59707	60945:	contig of 1239 bp in length	*
*		gap of unknown length		*
*	60946	61533:	contig of 588 bp in length	*
*		gap of unknown length		*
*	61534	62096:	contig of 563 bp in length	*
*		gap of unknown length		*
*	62097	62774:	contig of 678 bp in length	*
*		gap of unknown length		*
*	62775	63387:	contig of 613 bp in length	*
*		gap of unknown length		*
*	63388	63834:	contig of 447 bp in length	*
*		gap of unknown length		*
*	63835	64293:	contig of 459 bp in length	*
*		gap of unknown length		*
*	64294	64828:	contig of 535 bp in length	*
*		gap of unknown length		*
*	64829	65563:	contig of 735 bp in length	*
*		gap of unknown length		*
*	65564	65984:	contig of 421 bp in length	*
*		gap of unknown length		*
*	65985	66632:	contig of 648 bp in length	*
*		gap of unknown length		*
*	66633	67220:	contig of 588 bp in length	*
*		gap of unknown length		*
*	67221	67980:	contig of 760 bp in length	*
*		gap of unknown length		*
*	67981	68659:	contig of 679 bp in length	*
*		gap of unknown length		*
*	68660	69169:	contig of 510 bp in length	*
*		gap of unknown length		*
*	69170	69921:	gap of unknown length	*
*		contig of 752 bp in length		*
*	69922	70656:	gap of unknown length	*
*		contig of 735 bp in length		*
*	70657	71341:	gap of unknown length	*
*		contig of 685 bp in length		*
*	71342	72143:	gap of unknown length	*
*		contig of 802 bp in length		*
*	72144	72754:	gap of unknown length	*
*		contig of 611 bp in length		*
*	72755	73453:	gap of unknown length	*
*		contig of 699 bp in length		*
*	73454	74160:	gap of unknown length	*
*		contig of 707 bp in length		*
*	74161	74824:	gap of unknown length	*
*		contig of 664 bp in length		*
*	74825	75738:	gap of unknown length	*
*		contig of 914 bp in length		*
*	75739	76434:	gap of unknown length	*
*		contig of 696 bp in length		*
*	76435	76938:	gap of unknown length	*
*		contig of 504 bp in length		*
*	76939	77655:	gap of unknown length	*
*		contig of 717 bp in length		*
*	77656	78866:	gap of unknown length	*
*		contig of 1211 bp in length		*
*	78867	80003:	gap of unknown length	*
*		contig of 1137 bp in length		*
*	80004	80833:	gap of unknown length	*
*		contig of 830 bp in length		*
*	80834	81845:	gap of unknown length	*
*		contig of 1012 bp in length		*
*	81846	82882:	gap of unknown length	*
*		contig of 1037 bp in length		*
*	82883	83991:	gap of unknown length	*
*		contig of 1109 bp in length		*
*	83992	86061:	gap of unknown length	*
*		contig of 2070 bp in length		*
*	86062	86816:	gap of unknown length	*
*		contig of 755 bp in length		*
*	86817	88271:	gap of unknown length	*
*		contig of 1455 bp in length		*
*	88272	89289:	gap of unknown length	*
*		contig of 1018 bp in length		*
*	89290	90069:	gap of unknown length	*
*		contig of 780 bp in length		*
*	90070	91133:	gap of unknown length	*
*		contig of 1064 bp in length		*
*	91134	91833:	gap of unknown length	*
*		contig of 700 bp in length		*
*	91834	92661:	gap of unknown length	*
*		contig of 828 bp in length		*
*	92662	93414:	gap of unknown length	*
*		contig of 753 bp in length		*
*	93415	94471:	gap of unknown length	*
*		contig of 1057 bp in length		*
*	94472	95622:	gap of unknown length	*
*		contig of 1151 bp in length		*
*	95623	97279:	gap of unknown length	*
*		contig of 1657 bp in length		*
*	97280	98246:	gap of unknown length	*
*		contig of 967 bp in length		*
*	98247	99518:	gap of unknown length	*
*		contig of 1272 bp in length		*
*	99519	100837:	gap of unknown length	*
*		contig of 1319 bp in length		*
*	100838	102229:	gap of unknown length	*
*		contig of 1392 bp in length		*
*	102230	103965:	gap of unknown length	*
*		contig of 1736 bp in length		*
*	103966	105694:	gap of unknown length	*
*		contig of 1729 bp in length		*
*		gap of unknown length		*

```
Query Match      1.7%; Score 46; DB 41; Length 108064;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 17281 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17326

RESULT 20
AC017014.1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC017014 Accession AC017014
Fragment Name      Begin      End
AC017014_0         1      110000
AC017014_1        100001    210000
AC017014_2        200001    310000
AC017014_3        300001    367220
Continuation (2 of 4) of AC017014 from base 100001 (AC017014 Homo sapiens clone RP11-142)

Query Match      1.7%; Score 46; DB 45; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 34117 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 34072

RESULT 21
AC009889/c
LOCUS
DEFINITION
AC009889 132333 bp DNA HTG 04-SEP-1999
Homo sapiens clone H_NH0113K19, *** SEQUENCING IN PROGRESS ***, 13
unordered pieces.
AC009889
VERSION
AC009889.1 GI:5822964
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132333)
McCombie, W.R.
Human Genomic Sequence
Unpublished
2 (bases 1 to 132333)
McCombie, W.R.
Direct Submission
Submitted (04-SEP-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 21380: contig of 21380 bp in length
* 21381: 21885: gap of unknown length
* 21886: 40255: contig of 18570 bp in length
* 40256: 40360: gap of unknown length
* 40361: 57759: contig of 17199 bp in length
* 57760: 58064: gap of unknown length
* 58065: 72684: contig of 14620 bp in length
* 72685: 72989: gap of unknown length
* 72990: 87082: contig of 14093 bp in length
* 87083: 87387: gap of unknown length
* 87388: 98798: contig of 11411 bp in length
* 98799: 99102: gap of unknown length
```

```
* 99103 108656: contig of 9554 bp in length
* 108657 108960: gap of unknown length
* 108961 116080: contig of 7120 bp in length
* 116081 116384: gap of unknown length
* 116385 120735: contig of 4351 bp in length
* 120736 121039: gap of unknown length
* 121040 124805: contig of 3766 bp in length
* 124806 125110: gap of unknown length
* 125110 128610: contig of 3501 bp in length
* 128611 128914: gap of unknown length
* 128915 131764: contig of 2850 bp in length
* 131765 132068: gap of unknown length
* 132069 132333: contig of 265 bp in length.
FEATURES
Location/Qualifiers
1..132333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="H_NH0113K19"
BASE COUNT 37630 a 26434 c 26555 g 38040 t 3674 others
ORIGIN
Query Match      1.7%; Score 46; DB 41; Length 132333;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 131230 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 131185

RESULT 22
AL133259/c
LOCUS
DEFINITION
AL133259 140513 bp DNA HTG 30-NOV-1999
Homo sapiens chromosome 6 clone RP1-258E1, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
AL133259
VERSION
AL133259.1 GI:6522492
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140513)
Sims, S.
Direct Submission
Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00004 Length: 38566bp
Contig_ID: 00036 Length: 1729bp
Contig_ID: 00065 Length: 1361bp
Contig_ID: 00101 Length: 3279bp
Contig_ID: 00107 Length: 4463bp
Contig_ID: 00133 Length: 9012bp
Contig_ID: 00136 Length: 5087bp
Contig_ID: 00138 Length: 4932bp
Contig_ID: 00166 Length: 3138bp
Contig_ID: 00185 Length: 4561bp
Contig_ID: 00204 Length: 2642bp
Contig_ID: 00271 Length: 7384bp
Contig_ID: 00341 Length: 3289bp
Contig_ID: 00344 Length: 4762bp
Contig_ID: 00346 Length: 1650bp
Contig_ID: 00419 Length: 3620bp
Contig_ID: 00457 Length: 1422bp
Contig_ID: 00472 Length: 3681bp

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```

Contig_ID: 00515 Length: 2744bp
Contig_ID: 00520 Length: 2552bp
Contig_ID: 00530 Length: 1203bp
Contig_ID: 00601 Length: 8631bp
Contig_ID: 00644 Length: 4896bp
Contig_ID: 00662 Length: 7143bp
Contig_ID: 00679 Length: 1754bp
Contig_ID: 00697 Length: 1162bp
Contig_ID: 00700 Length: 1211bp
Contig_ID: 00850 Length: 3517bp
Contig_ID: 00867 Length: 2919bp
Contig_ID: 00878 Length: 1163bp
Contig_ID: 00894 Length: 5149bp
Contig_ID: 00912 Length: 1801bp
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES             Location/Qualifiers
     source            1..140513
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="q24-25"
                        /clone="RP1-258E1"
                        /clone_lib="RPCI-1"
BASE COUNT      32737 a 24991 c 24342 g 33637 t 24806 others
ORIGIN

Query Match          1.7%; Score 46; DB 33; Length 140513;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
        |||||...|||
Db 119775 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 119730

RESULT 23
HSDJ315G1          HSDJ315G1 144387 bp DNA HTG 27-NOV-1999
LOCUS              Homo sapiens chromosome X clone RP1-315G1 map q24-25, ***
DEFINITION          SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION            ALI21601.3 GI:6478115
VERSION              ALI21601.3 GI:6478115
KEYWORDS             HTG; HTGS_PHASE1.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (bases 1 to 144387)
AUTHORS              Pavitt, R.
TITLE                Direct Submission
JOURNAL              Submitted (27-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
                    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT              On Nov 29, 1999 this sequence is unfinished and does not necessarily
                    represent the correct sequence. Work on the sequence is in
                    progress and the release of this data is based on the understanding
                    that the sequence may change as work continues. The sequence may
                    be contaminated with foreign sequence from E.coli, yeast, vector,
                    phage etc. Order of segments is not known; 800 n's separate
                    segments. Contig_ID: 00024 Length: 1179bp
                    Contig_ID: 00226 Length: 1287bp
                    Contig_ID: 00281 Length: 7809bp
                    Contig_ID: 00308 Length: 10670bp
                    Contig_ID: 00334 Length: 23722bp
                    Contig_ID: 00341 Length: 5230bp
                    Contig_ID: 00491 Length: 6598bp
                    Contig_ID: 00648 Length: 6405bp
                    Contig_ID: 00949 Length: 26471bp
                    Contig_ID: 01110 Length: 5629bp

FEATURES             Location/Qualifiers
     source            1..140513
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="q24-25"
                        /clone="RP1-258E1"
                        /clone_lib="RPCI-1"
BASE COUNT      32737 a 24991 c 24342 g 33637 t 24806 others
ORIGIN

Query Match          1.7%; Score 46; DB 33; Length 140513;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
        |||||...|||
Db 119775 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 119730

RESULT 23
HSDJ315G1          HSDJ315G1 144387 bp DNA HTG 27-NOV-1999
LOCUS              Homo sapiens chromosome X clone RP1-315G1 map q24-25, ***
DEFINITION          SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION            ALI21601.3 GI:6478115
VERSION              ALI21601.3 GI:6478115
KEYWORDS             HTG; HTGS_PHASE1.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (bases 1 to 144387)
AUTHORS              Pavitt, R.
TITLE                Direct Submission
JOURNAL              Submitted (27-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
                    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT              On Nov 29, 1999 this sequence is unfinished and does not necessarily
                    represent the correct sequence. Work on the sequence is in
                    progress and the release of this data is based on the understanding
                    that the sequence may change as work continues. The sequence may
                    be contaminated with foreign sequence from E.coli, yeast, vector,
                    phage etc. Order of segments is not known; 800 n's separate
                    segments. Contig_ID: 00024 Length: 1179bp
                    Contig_ID: 00226 Length: 1287bp
                    Contig_ID: 00281 Length: 7809bp
                    Contig_ID: 00308 Length: 10670bp
                    Contig_ID: 00334 Length: 23722bp
                    Contig_ID: 00341 Length: 5230bp
                    Contig_ID: 00491 Length: 6598bp
                    Contig_ID: 00648 Length: 6405bp
                    Contig_ID: 00949 Length: 26471bp
                    Contig_ID: 01110 Length: 5629bp

FEATURES             Location/Qualifiers
     source            1..140513
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /map="q24-25"
                        /clone="RP1-258E1"
                        /clone_lib="RPCI-1"
BASE COUNT      32737 a 24991 c 24342 g 33637 t 24806 others
ORIGIN

Query Match          1.7%; Score 46; DB 33; Length 140513;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
        |||||...|||
Db 119775 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 119730

RESULT 24
AC010129           AC010129 163195 bp DNA HTG 05-NOV-1999
LOCUS              Homo sapiens clone NH0059N09, *** SEQUENCING IN PROGRESS ***, 2
DEFINITION          unordered pieces.
ACCESSION            AC010129
VERSION              AC010129.2 GI:6249775
KEYWORDS             HTG; HTGS_PHASE1.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (bases 1 to 163195)
AUTHORS              Waterston, R.H.
TITLE                The sequence of Homo sapiens clone
JOURNAL              Unpublished
                    Waterston, R.H.
REFERENCE            2 (bases 1 to 163195)
AUTHORS              Waterston, R.H.
TITLE                Direct Submission
JOURNAL              Submitted (13-SEP-1999) Genome Sequencing Center, Washington
                    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                    MO 63108, USA
COMMENT              On Nov 5, 1999 this sequence version replaced gi:5871004.
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 2 contigs. The true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.
                    * 1 48775: contig of 48775 bp in length
                    * 48776 48793: gap of unknown length
                    * 48794 163195: contig of 114402 bp in length.

FEATURES             Location/Qualifiers
     source            1..163195
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="NH0059N09"
BASE COUNT      52860 a 30571 c 29430 g 50296 t 18 others
ORIGIN

Query Match          1.7%; Score 46; DB 43; Length 163195;

```

```

Contig_ID: 01138 Length: 5229bp
Contig_ID: 01398 Length: 9434bp
Contig_ID: 01430 Length: 1563bp
Contig_ID: 01776 Length: 22761bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES             Location/Qualifiers
     source            1..144387
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="X"
                        /map="q24-25"
                        /clone="RP1-315G1"
                        /clone_lib="RPCI-1"
BASE COUNT      40957 a 28225 c 27663 g 37139 t 10403 others
ORIGIN

Query Match          1.7%; Score 46; DB 33; Length 144387;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
        |||||...|||
Db 46903 CCCCCAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 46948

RESULT 24
AC010129           AC010129 163195 bp DNA HTG 05-NOV-1999
LOCUS              Homo sapiens clone NH0059N09, *** SEQUENCING IN PROGRESS ***, 2
DEFINITION          unordered pieces.
ACCESSION            AC010129
VERSION              AC010129.2 GI:6249775
KEYWORDS             HTG; HTGS_PHASE1.
SOURCE               human.
ORGANISM             Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (bases 1 to 163195)
AUTHORS              Waterston, R.H.
TITLE                The sequence of Homo sapiens clone
JOURNAL              Unpublished
                    Waterston, R.H.
REFERENCE            2 (bases 1 to 163195)
AUTHORS              Waterston, R.H.
TITLE                Direct Submission
JOURNAL              Submitted (13-SEP-1999) Genome Sequencing Center, Washington
                    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                    MO 63108, USA
COMMENT              On Nov 5, 1999 this sequence version replaced gi:5871004.
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 2 contigs. The true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.
                    * 1 48775: contig of 48775 bp in length
                    * 48776 48793: gap of unknown length
                    * 48794 163195: contig of 114402 bp in length.

FEATURES             Location/Qualifiers
     source            1..163195
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="NH0059N09"
BASE COUNT      52860 a 30571 c 29430 g 50296 t 18 others
ORIGIN

Query Match          1.7%; Score 46; DB 43; Length 163195;

```

Best Local Similarity 100.0%; Pred. No. 2.5e-14; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAA 2689
|||||
Db 568 CCCCCAAA 713
|||||

RESULT 25
AC011280
LOCUS Homo sapiens clone 13_K_8, *** SEQUENCING IN PROGRESS ***, 3
DEFINITION AC011280 163533 bp DNA HTG 05-NOV-1999
unordered pieces.
ACCESSION AC011280
VERSION AC011280.2 GI:6249712
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163533)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone 13_K_8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163533)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Casare,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEvan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 5, 1999 this sequence version replaced gi:601210.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20081: contig of 20081 bp in length
* gap of unknown length
* 20082 50254: contig of 30173 bp in length
* gap of unknown length
* 50255 163533: contig of 113279 bp in length.
FEATURES
Location/Qualifiers
1..163533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="13_K_8"
/clone_lib="RPC1-11 Human Male BAC"

BASE COUNT 53421 a 30093 c 29840 g 50177 t 2 others
ORIGIN

Query Match 1.7%; Score 46; DB 43; Length 163533;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAA 2689
|||||
Db 20431 CCCCCAAA 20476
|||||

RESULT 26
AC012067
LOCUS Homo sapiens clone NH0192N14, *** SEQUENCING IN PROGRESS ***, 4
DEFINITION AC012067 165284 bp DNA HTG 19-OCT-1999
unordered pieces.
ACCESSION AC012067
VERSION AC012067.1 GI:6067210
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 165284)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165284)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8528: contig of 8528 bp in length
* 8529 8545: gap of unknown length
* 8546 49205: contig of 40660 bp in length
* 49206 49222: gap of unknown length
* 49223 99112: contig of 49890 bp in length
* 99113 99129: gap of unknown length
* 99130 165284: contig of 66155 bp in length.
FEATURES
Location/Qualifiers
1..165284
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0192N14"

BASE COUNT 53579 a 29550 c 29161 g 52943 t 51 others
ORIGIN

Query Match 1.7%; Score 46; DB 42; Length 165284;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAA 2689
|||||
Db 413 CCCCCAAA 458
|||||

RESULT 27
AC012457/c
LOCUS Homo sapiens clone NH0509E16, *** SEQUENCING IN PROGRESS ***, 17
DEFINITION AC012457 165414 bp DNA HTG 27-OCT-1999
unordered pieces.
ACCESSION AC012457
VERSION AC012457.1 GI:6136429
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 165414)


```
repeat_region 6746..6796 /rpt_family="MERL_type"
repeat_region 7119..7216 /rpt_family="MaLR"
repeat_region 7373..7440 /rpt_family="MERL_type"
repeat_region 8177..8306 /rpt_family="L2"
repeat_region 8919..9756 /rpt_family="L1"
repeat_region 9799..11055 /rpt_family="L1"
repeat_region 11062..11096 /rpt_family="(GTTG)n"
repeat_region 11097..11374 /rpt_family="Alu"
repeat_region 11375..12677 /rpt_family="L1"
repeat_region 13272..13558 /rpt_family="Alu"
repeat_region 13954..14036 /rpt_family="L1"
repeat_region 14037..14069 /rpt_family="AT-rich"
repeat_region 14699..14821 /rpt_family="(TA)n"
repeat_region 15009..15137 /rpt_family="(TA)n"
repeat_region 15110..15459 /rpt_family="(TATATG)n"
repeat_region 15463..15565 /rpt_family="(TA)n"
repeat_region 15568..15741 /rpt_family="(TATAA)n"
repeat_region 15755..15915 /rpt_family="(TA)n"
repeat_region 16692..17574 /rpt_family="L1"
repeat_region 17575..17827 /rpt_family="Alu"
repeat_region 17828..18595 /rpt_family="L1"
repeat_region 18596..19002 /rpt_family="MaLR"
repeat_region 19003..20453 /rpt_family="L1"
repeat_region 20484..20801 /rpt_family="L1"
repeat_region 20802..20829 /rpt_family="L1"
repeat_region 20830..21249 /rpt_family="(TAAAA)n"
repeat_region 21250..21545 /rpt_family="Alu"
repeat_region 21546..22591 /rpt_family="L1"
repeat_region 22601..23607 /rpt_family="L1"
repeat_region 23608..23696 /rpt_family="Alu"
repeat_region 23698..24789 /rpt_family="L1"
repeat_region 24790..24824 /rpt_family="(CA)n"
repeat_region 24825..25046 /rpt_family="L1"
repeat_region 26174..26540 /rpt_family="MaLR"
repeat_region 26601..28221 /rpt_family="MaLR"
repeat_region 28222..28520 /rpt_family="MaLR"
repeat_region 29411..29442
```

```
repeat_region /rpt_family="(TCCC)n"
misc_feature 29438..29479 /rpt_family="CT-rich"
misc_feature 30423..30503 /note="pseudogene similar to PID:g3970874 (AB015344)
HRIHFB2157 [Homo sapiens]"
misc_feature 30506..31558 /note="pseudogene similar to PID:g3970874 (AB015344)
HRIHFB2157 [Homo sapiens]"
exon 31819..32103
repeat_region 32277..33564 /rpt_family="L2"
repeat_region 33838..34006 /rpt_family="MIR"
repeat_region 34246..34583 /rpt_family="MaLR"
repeat_region 34596..34784 /rpt_family="MERL_type"
repeat_region 35006..35126 /rpt_family="MERL_type"
repeat_region 35127..35247 /rpt_family="Alu"
repeat_region 35248..35315 /rpt_family="MERL_type"
repeat_region 35519..35668 /rpt_family="MIR"
repeat_region 35794..36077 /rpt_family="Alu"
repeat_region 36286..36497 /rpt_family="L2"
repeat_region 37034..37263 /rpt_family="Alu"
repeat_region 37655..37837 /rpt_family="Retroviral"
repeat_region 39386..39557 /rpt_family="MIR"
repeat_region 39639..39760 /rpt_family="L2"
repeat_region 39794..39864 /rpt_family="L2"
repeat_region 40988..41012 /rpt_family="(T)n"
repeat_region 41091..41348 /rpt_family="MIR"
repeat_region 41721..42106 /rpt_family="Retroviral"
repeat_region 43882..43999 /rpt_family="L2"
repeat_region 44623..44669 /rpt_family="(TG)n"
repeat_region 44756..44953
```

Query Match 1.7%; Score 46; DB 42; Length 155445;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Db 161469 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 161424

RESULT 29
AC012114/c AC012114 168363 bp DNA HTG 20-OCT-1999
LOCUS Homo sapiens clone l_H_12, *** SEQUENCING IN PROGRESS ***, 25
DEFINITION unordered pieces.
ACCESSION AC012114
VERSION AC012114.1 GI:6088012
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

1 (bases 1 to 168363)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Homo sapiens, clone 1_H_12
Unpublished
REFERENCE
JOURNAL
AUTHORS
1 (bases 1 to 168363)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castie,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1186: gap of unknown length
2 1187: contig of 1186 bp in length
3 1188: gap of unknown length
4 1189: contig of 1186 bp in length
5 1190: gap of unknown length
6 1191: contig of 1186 bp in length
7 1192: gap of unknown length
8 1193: contig of 1186 bp in length
9 1194: gap of unknown length
10 1195: contig of 1186 bp in length
11 1196: gap of unknown length
12 1197: contig of 1186 bp in length
13 1198: gap of unknown length
14 1199: contig of 1186 bp in length
15 1200: gap of unknown length
16 1201: contig of 1186 bp in length
17 1202: gap of unknown length
18 1203: contig of 1186 bp in length
19 1204: gap of unknown length
20 1205: contig of 1186 bp in length
21 1206: gap of unknown length
22 1207: contig of 1186 bp in length
23 1208: gap of unknown length
24 1209: contig of 1186 bp in length
25 1210: gap of unknown length
26 1211: contig of 1186 bp in length
27 1212: gap of unknown length
28 1213: contig of 1186 bp in length
29 1214: gap of unknown length
30 1215: contig of 1186 bp in length
31 1216: gap of unknown length
32 1217: contig of 1186 bp in length
33 1218: gap of unknown length
34 1219: contig of 1186 bp in length
35 1220: gap of unknown length
36 1221: contig of 1186 bp in length
37 1222: gap of unknown length
38 1223: contig of 1186 bp in length
39 1224: gap of unknown length
40 1225: contig of 1186 bp in length
41 1226: gap of unknown length
42 1227: contig of 1186 bp in length
43 1228: gap of unknown length
44 1229: contig of 1186 bp in length
45 1230: gap of unknown length
46 1231: contig of 1186 bp in length
47 1232: gap of unknown length
48 1233: contig of 1186 bp in length
49 1234: gap of unknown length
50 1235: contig of 1186 bp in length
51 1236: gap of unknown length
52 1237: contig of 1186 bp in length
53 1238: gap of unknown length
54 1239: contig of 1186 bp in length
55 1240: gap of unknown length
56 1241: contig of 1186 bp in length
57 1242: gap of unknown length
58 1243: contig of 1186 bp in length
59 1244: gap of unknown length
60 1245: contig of 1186 bp in length
61 1246: gap of unknown length
62 1247: contig of 1186 bp in length
63 1248: gap of unknown length
64 1249: contig of 1186 bp in length
65 1250: gap of unknown length
66 1251: contig of 1186 bp in length
67 1252: gap of unknown length
68 1253: contig of 1186 bp in length
69 1254: gap of unknown length
70 1255: contig of 1186 bp in length
71 1256: gap of unknown length
72 1257: contig of 1186 bp in length
73 1258: gap of unknown length
74 1259: contig of 1186 bp in length
75 1260: gap of unknown length
76 1261: contig of 1186 bp in length
77 1262: gap of unknown length
78 1263: contig of 1186 bp in length
79 1264: gap of unknown length
80 1265: contig of 1186 bp in length
81 1266: gap of unknown length
82 1267: contig of 1186 bp in length
83 1268: gap of unknown length
84 1269: contig of 1186 bp in length
85 1270: gap of unknown length
86 1271: contig of 1186 bp in length
87 1272: gap of unknown length
88 1273: contig of 1186 bp in length
89 1274: gap of unknown length
90 1275: contig of 1186 bp in length
91 1276: gap of unknown length
92 1277: contig of 1186 bp in length
93 1278: gap of unknown length
94 1279: contig of 1186 bp in length
95 1280: gap of unknown length
96 1281: contig of 1186 bp in length
97 1282: gap of unknown length
98 1283: contig of 1186 bp in length
99 1284: gap of unknown length
100 1285: contig of 1186 bp in length
101 1286: gap of unknown length
102 1287: contig of 1186 bp in length
103 1288: gap of unknown length
104 1289: contig of 1186 bp in length
105 1290: gap of unknown length
106 1291: contig of 1186 bp in length
107 1292: gap of unknown length
108 1293: contig of 1186 bp in length
109 1294: gap of unknown length
110 1295: contig of 1186 bp in length
111 1296: gap of unknown length
112 1297: contig of 1186 bp in length
113 1298: gap of unknown length
114 1299: contig of 1186 bp in length
115 1300: gap of unknown length
116 1301: contig of 1186 bp in length
117 1302: gap of unknown length
118 1303: contig of 1186 bp in length
119 1304: gap of unknown length
120 1305: contig of 1186 bp in length
121 1306: gap of unknown length
122 1307: contig of 1186 bp in length
123 1308: gap of unknown length
124 1309: contig of 1186 bp in length
125 1310: gap of unknown length
126 1311: contig of 1186 bp in length
127 1312: gap of unknown length
128 1313: contig of 1186 bp in length
129 1314: gap of unknown length
130 1315: contig of 1186 bp in length
131 1316: gap of unknown length
132 1317: contig of 1186 bp in length
133 1318: gap of unknown length
134 1319: contig of 1186 bp in length
135 1320: gap of unknown length
136 1321: contig of 1186 bp in length
137 1322: gap of unknown length
138 1323: contig of 1186 bp in length
139 1324: gap of unknown length
140 1325: contig of 1186 bp in length
141 1326: gap of unknown length
142 1327: contig of 1186 bp in length
143 1328: gap of unknown length
144 1329: contig of 1186 bp in length
145 1330: gap of unknown length
146 1331: contig of 1186 bp in length
147 1332: gap of unknown length
148 1333: contig of 1186 bp in length
149 1334: gap of unknown length
150 1335: contig of 1186 bp in length
151 1336: gap of unknown length
152 1337: contig of 1186 bp in length
153 1338: gap of unknown length
154 1339: contig of 1186 bp in length
155 1340: gap of unknown length
156 1341: contig of 1186 bp in length
157 1342: gap of unknown length
158 1343: contig of 1186 bp in length
159 1344: gap of unknown length
160 1345: contig of 1186 bp in length
161 1346: gap of unknown length
162 1347: contig of 1186 bp in length
163 1348: gap of unknown length
164 1349: contig of 1186 bp in length
165 1350: gap of unknown length
166 1351: contig of 1186 bp in length
167 1352: gap of unknown length
168 1353: contig of 1186 bp in length
169 1354: gap of unknown length
170 1355: contig of 1186 bp in length
171 1356: gap of unknown length
172 1357: contig of 1186 bp in length
173 1358: gap of unknown length
174 1359: contig of 1186 bp in length
175 1360: gap of unknown length
176 1361: contig of 1186 bp in length
177 1362: gap of unknown length
178 1363: contig of 1186 bp in length
179 1364: gap of unknown length
180 1365: contig of 1186 bp in length
181 1366: gap of unknown length
182 1367: contig of 1186 bp in length
183 1368: gap of unknown length
184 1369: contig of 1186 bp in length
185 1370: gap of unknown length
186 1371: contig of 1186 bp in length
187 1372: gap of unknown length
188 1373: contig of 1186 bp in length
189 1374: gap of unknown length
190 1375: contig of 1186 bp in length
191 1376: gap of unknown length
192 1377: contig of 1186 bp in length
193 1378: gap of unknown length
194 1379: contig of 1186 bp in length
195 1380: gap of unknown length
196 1381: contig of 1186 bp in length
197 1382: gap of unknown length
198 1383: contig of 1186 bp in length
199 1384: gap of unknown length
200 1385: contig of 1186 bp in length
201 1386: gap of unknown length
202 1387: contig of 1186 bp in length
203 1388: gap of unknown length
204 1389: contig of 1186 bp in length
205 1390: gap of unknown
```

```
Center project name: L3777
Center clone name: 363.G.3
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142869 bases at least Q40
Consensus quality: 157748 bases at least Q30
Consensus quality: 164454 bases at least Q20
Insert size: 169168; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 675: contig of 675 bp in length
*
* 676 2714: contig of 2039 bp in length
*
* 2715 5561: contig of 2847 bp in length
*
* 5562 7526: contig of 1965 bp in length
*
* 7527 9378: contig of 1852 bp in length
*
* 9379 19384: contig of 10006 bp in length
*
* 19385 26444: contig of 7060 bp in length
*
* 26445 35340: contig of 8896 bp in length
*
* 35341 44797: contig of 9457 bp in length
*
* 44798 54146: contig of 9349 bp in length
*
* 54147 64634: contig of 10488 bp in length
*
* 64635 76415: contig of 11781 bp in length
*
* 76416 97235: contig of 20820 bp in length
*
* 97236 129656: contig of 32421 bp in length
*
* 129657 169168: contig of 39512 bp in length.
*
* Location/Qualifiers
* 1..169168
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="2"
* /map="2"
* /clone="RP11-363G3"
* /clone_lib="RPCI-11 Human Male BAC"
*
* 54711 a 28671 c 28866 g 56914 t 6 others
*
* BASE COUNT 54711 a 28671 c 28866 g 56914 t
* ORIGIN
*
* Query Match 1.7%; Score 46; DB 44; Length 169168;
* Best Local Similarity 100.0%; Pred. No. 2.5e-14;
* Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
* | | | | | | | | | | | | | | | | | | | | | | | | | | | |
* Db 19636 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 19681
*
* REFERENCE
* 1 (bases 1 to 169168)
* Kohalmi, S.E., Schorr, P., Nowak, J., Reader, L.J. and Crosby, W.L.
* The Product of AGAMOUS Interacts with Select Members of the
* MADS-Box Family of Proteins in Arabidopsis thaliana
* Unpublished
* 2 (bases 1 to 1016)
* Kohalmi, S.E., Schorr, P., Nowak, J., Reader, L.J. and Crosby, W.L.
* Direct Submission
* Submitted (22-MAY-1997) Gene Expression Group, NRC Plant
* Biotechnology Institute, 110 Gymnasium Place, Saskatoon, SK S7N
* 0W9, Canada
*
* RESULT 31
* S5Y10781
* LOCUS 342 bp mRNA PLN 14-OCT-1998
```

```
DEFINITION S.stapfianus pSD.8b mRNA.
ACCESSION Y10781
VERSION Y10781.1 GI:1808681
KEYWORDS desiccation tolerance.
SOURCE Sporobolus stapfianus.
ORGANISM Sporobolus stapfianus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Sporobolus.
REFERENCE 1 (bases 1 to 342)
Blomstedt, C.K., Gianello, R.D., Neale, A.D., Hamill, J.D. and
Gaif, D.P.
Isolation and characterization of cDNAs associated with the onset
of desiccation tolerance in the resurrection grass Sporobolus
stapfianus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 342)
Blomstedt, C.K.
Direct Submission
TITLE Submitted (20-JAN-1997) C.K. Blomstedt, Monash University, Genetics
And Developmental Biology, Wellington Road, Clayton, Melbourne,
Victoria 3168, AUSTRALIA
JOURNAL
FEATURES
source
1..342
/organism="Sporobolus stapfianus"
/variety="Gandoger"
/db_xref="taxon:56623"
/tissue_type="leaf"
/clone="pSD.8b"
<1..60
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA71753.1"
/db_xref="GI:1808682"
/db_xref="SPTREMBL:O04817"
/db_xref="translation="OKLYDDLTSGERNLEDEAR"
BASE COUNT 127 a 59 c 67 g 89 t
ORIGIN
*
* Query Match 1.7%; Score 45; DB 7; Length 342;
* Best Local Similarity 100.0%; Pred. No. 9.3e-14;
* Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2688
* | | | | | | | | | | | | | | | | | | | | | | | | | | | |
* Db 298 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 342
*
* RESULT 32
* AF005158
* LOCUS 1016 bp mRNA PLN 07-OCT-1998
* DEFINITION Arabidopsis thaliana MADS-box Protein (AGL24) mRNA, complete cds.
* ACCESSION AF005158
* VERSION AF005158.1 GI:3719214
* KEYWORDS
* SOURCE thale cress.
* ORGANISM Arabidopsis thaliana
* Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
* euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
* eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
* Arabidopsis.
* REFERENCE 1 (bases 1 to 1016)
* Kohalmi, S.E., Schorr, P., Nowak, J., Reader, L.J. and Crosby, W.L.
* The Product of AGAMOUS Interacts with Select Members of the
* MADS-Box Family of Proteins in Arabidopsis thaliana
* Unpublished
* 2 (bases 1 to 1016)
* Kohalmi, S.E., Schorr, P., Nowak, J., Reader, L.J. and Crosby, W.L.
* Direct Submission
* Submitted (22-MAY-1997) Gene Expression Group, NRC Plant
* Biotechnology Institute, 110 Gymnasium Place, Saskatoon, SK S7N
* 0W9, Canada
```

```
Query Match      1.7%; Score 45; DB 10; Length 3105;
Best Local Similarity 100.0%; Pred. No. 9.2e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


BASE COUNT
ORIGIN
RCYRKLHLSPTDPTVWLRLRSVDRSLLOGWODGFIPTANVVLYMLCRDVISSEVGS
DHELOAVLITCLVLSYSYMGNEISYPLAFVESCCKEAFWDRCLSVINLMSSKMLQIN
ADPHFTQVFSDLKNESQOEKKRLLLGLDR
694 a 1113 c 1084 g 691 t

Query Match 1-7%; Score 45; DB 3; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.2e-14;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||

Db 3538 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3582

RESULT 38

AC008696/c

LOCUS 41684 bp DNA HTG 03-AUG-1999

DEFINITION Homo sapiens chromosome 5 clone CIT978SKB_70D19, *** SEQUENCING IN

PROGRESS ***; 83 unordered pieces.

AC008696

VERSION AC008696.1 GI:5686375

KEYWORDS HTG: HTGS_PHASE1.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 41684)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 5

Unpublished

2 (bases 1 to 41684)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

www.jgi.doe.gov.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 83 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 661: contig of 661 bp in length

gap of unknown length

662 1321: contig of 660 bp in length

gap of unknown length

1322 1605: contig of 284 bp in length

gap of unknown length

1606 2158: contig of 553 bp in length

gap of unknown length

2159 2814: contig of 656 bp in length

gap of unknown length

2815 2913: contig of 99 bp in length

gap of unknown length

2914 2976: contig of 63 bp in length

gap of unknown length

2977 3046: contig of 70 bp in length

gap of unknown length

3047 4099: contig of 1053 bp in length

gap of unknown length

4100 5221: contig of 1122 bp in length

gap of unknown length

5222 5877: contig of 656 bp in length

gap of unknown length

5878 6061: contig of 184 bp in length

gap of unknown length

6062 6178: contig of 117 bp in length

gap of unknown length

6179 6636: contig of 458 bp in length

6637 6978: contig of 242 bp in length
gap of unknown length
6879 6940: contig of 62 bp in length
gap of unknown length
6941 7030: contig of 90 bp in length
gap of unknown length
7031 8041: contig of 1011 bp in length
gap of unknown length
8042 8098: contig of 57 bp in length
gap of unknown length
8099 8423: contig of 325 bp in length
gap of unknown length
8424 9390: contig of 967 bp in length
gap of unknown length
9391 9978: contig of 588 bp in length
gap of unknown length
9979 10631: contig of 653 bp in length
gap of unknown length
10632 11504: contig of 873 bp in length
gap of unknown length
11505 11550: contig of 45 bp in length
gap of unknown length
11551 11779: contig of 229 bp in length
gap of unknown length
11780 11870: contig of 91 bp in length
gap of unknown length
11871 11941: contig of 71 bp in length
gap of unknown length
11942 12029: contig of 88 bp in length
gap of unknown length
12030 12122: contig of 93 bp in length
gap of unknown length
12123 12224: contig of 102 bp in length
gap of unknown length
12225 13275: contig of 1051 bp in length
gap of unknown length
13276 13390: contig of 115 bp in length
gap of unknown length
13391 14045: contig of 655 bp in length
gap of unknown length
14046 14963: contig of 918 bp in length
gap of unknown length
14964 15144: contig of 181 bp in length
gap of unknown length
15145 15812: contig of 668 bp in length
gap of unknown length
15813 15889: contig of 77 bp in length
gap of unknown length
15890 16083: contig of 194 bp in length
gap of unknown length
16084 16175: contig of 92 bp in length
gap of unknown length
16176 16773: contig of 598 bp in length
gap of unknown length
16774 17972: contig of 1199 bp in length
gap of unknown length
17973 18078: contig of 106 bp in length
gap of unknown length
18079 19129: contig of 1051 bp in length
gap of unknown length
19130 19690: contig of 561 bp in length
gap of unknown length
19691 19772: contig of 82 bp in length
gap of unknown length
19773 20041: contig of 269 bp in length
gap of unknown length
20042 20398: contig of 357 bp in length
gap of unknown length
20399 21269: contig of 871 bp in length
gap of unknown length
21270 21399: contig of 130 bp in length
gap of unknown length

```

* 21400 22156: contig of 757 bp in length
*      gap of unknown length
* 22157 22224: contig of 68 bp in length
*      gap of unknown length
* 22225 22868: contig of 644 bp in length
*      gap of unknown length
* 22869 22992: contig of 124 bp in length
*      gap of unknown length
* 22993 23626: contig of 634 bp in length
*      gap of unknown length
* 23627 24151: contig of 525 bp in length
*      gap of unknown length
* 24152 24226: contig of 75 bp in length
*      gap of unknown length
* 24227 25033: contig of 807 bp in length
*      gap of unknown length
* 25034 25084: contig of 51 bp in length
*      gap of unknown length
* 25085 25269: contig of 185 bp in length
*      gap of unknown length
* 25270 25427: contig of 158 bp in length
*      gap of unknown length
* 25428 25985: contig of 558 bp in length
*      gap of unknown length
* 25986 26620: contig of 635 bp in length
*      gap of unknown length
* 26621 27493: contig of 873 bp in length
*      gap of unknown length
* 27494 27654: contig of 161 bp in length
*      gap of unknown length
* 27655 27762: contig of 108 bp in length
*      gap of unknown length
* 27763 27993: contig of 231 bp in length
*      gap of unknown length
* 27994 28208: contig of 215 bp in length
*      gap of unknown length
* 28209 28825: contig of 617 bp in length
*      gap of unknown length
* 28826 28908: contig of 83 bp in length
*      gap of unknown length
* 28909 29036: contig of 128 bp in length
*      gap of unknown length
* 29037 30130: contig of 1094 bp in length
*      gap of unknown length
* 30131 31238: contig of 1108 bp in length
*      gap of unknown length
* 31239 31788: contig of 550 bp in length
*      gap of unknown length
* 31789 32692: contig of 904 bp in length
*      gap of unknown length
* 32693 34522: contig of 1830 bp in length
*      gap of unknown length
* 34523 35721: contig of 1199 bp in length
*      gap of unknown length
* 35722 37271: contig of 1550 bp in length
*      gap of unknown length
* 37272 37665: contig of 394 bp in length
*      gap of unknown length
* 37666 38846: contig of 1181 bp in length
*      gap of unknown length
* 38847 39901: contig of 1055 bp in length
*      gap of unknown length
* 39902 41076: contig of 1175 bp in length
*      gap of unknown length
* 41077 41684: contig of 608 bp in length.
*      Location/Qualifiers
*      1. .41684
*         /organism="Homo sapiens"
*         /db_xref="taxon:9606"
*         /chromosome="5"
*         /clone="CIT978SKB_70D19"
BASE COUNT 11911 a 8751 c 8789 g 12204 t 29 others

```

```

Query Match      1.7%; Score 45; DB 41; Length 41684;
Best Local Similarity 100.0%; Pred. No. 9.1e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2645 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC|
Db 13880 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 13836

RESULT 39
AC008251/c      55868 bp      DNA      HTG      02-AUG-1999
Drosophila melanogaster chromosome 3 clone DS00111 (b456) map
65D2-65D3 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 47
unordered pieces.

AC008251      AC008251.1 GI:5656706
VERSION      HTG; HTGS_PHASE1.
KEYWORDS      fruit fly.
SOURCE
ORGANISM      Drosophila melanogaster
REFERENCE      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS      Butenihoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
      Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
      Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
      Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
      Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
      Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
      Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
      Rubin,G.M.
TITLE      Sequencing of Drosophila melanogaster
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 55868)
AUTHORS      Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
      Butenihoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
      Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
      Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
      Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
      Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
      Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
      Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
      Rubin,G.M.
TITLE      Direct Submission
JOURNAL      Submitted (30-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
COMMENT      Laboratory, MS 64-121, Berkeley, CA 94720, USA
      For further information about this sequence, including its location
      and relationship to other sequences, please visit our sequence
      archive web site (http://www.fruitfly.org/sequence/) or send email
      to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
      the following cutoffs: length >= 200 bases. Pl library location:
      2-15.
      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 47 contigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      * 1      749: contig of 749 bp in length
      *      829: gap of unknown length
      *      830      1485: contig of 656 bp in length
      *      1486      1565: gap of unknown length
      *      1566      2255: contig of 690 bp in length
      *      2256      2335: gap of unknown length
      *      2336      3040: contig of 705 bp in length
      *      3041      3120: gap of unknown length
      *      3121      3781: contig of 661 bp in length
      *      3782      3861: gap of unknown length
      *      3862      4420: contig of 559 bp in length

```



```

FEATURES
  source
    Location/Qualifiers
      1. 111079
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /map="q31"
        /clone_lib="RPCI-3"
        /clone="RP3-419C19"
BASE COUNT 35388 a 21441 c 21433 g 32817 t
ORIGIN

Query Match 1.7%; Score 45; DB 11; Length 111079;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
DB 13113 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 13069

RESULT 41
AC011414/c
LOCUS Homo sapiens chromosome 5 clone CIT978SKB_83D3, *** SEQUENCING IN
DEFINITION PROGRESS ***, 11 unordered pieces.
ACCESSION AC011414
VERSION AC011414.1 GI:6013528
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 125260)
TITLE DOE Joint Genome Institute.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 125260)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 660: contig of 660 bp in length
gap of unknown length
661 1339: contig of 679 bp in length
gap of unknown length
1340 2021: contig of 682 bp in length
gap of unknown length
2022 5034: gap of unknown length
gap of 3013 bp in length
5035 10078: contig of 5044 bp in length
gap of unknown length
10079 21131: contig of 11053 bp in length
gap of unknown length
21132 33664: contig of 12533 bp in length
gap of unknown length
33665 46169: contig of 12505 bp in length
gap of unknown length
46170 63481: contig of 17312 bp in length
gap of unknown length
63482 92267: contig of 28786 bp in length
gap of unknown length
92268 125260: contig of 32993 bp in length.

```

```

FEATURES
  source
    Location/Qualifiers
      1. 125260
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="CIT978SKB_83D3"
BASE COUNT 40541 a 24231 c 24261 g 36163 t
ORIGIN

Query Match 1.7%; Score 45; DB 41; Length 125260;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
DB 11448 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11404

RESULT 42
HSDJ319M7/c
LOCUS Homo sapiens chromosome 6 clone RPI-319M7 map p21.1-21.3, ***
DEFINITION SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AL079341
VERSION AL079341.14 GI:6434655
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 128208)
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Nov 15, 1999 this sequence version replaced gi:6136977.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 01957 Length: 128208bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1. 128208
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPI-319M7"
/clone_lib="RPCI-1"
/map="p21.1-21.3"
BASE COUNT 36143 a 23938 c 26492 g 41630 t
ORIGIN

Query Match 1.7%; Score 45; DB 33; Length 128208;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
DB 101579 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 101535

RESULT 43
AC011262
LOCUS AC011262 145346 bp DNA HTG 19-NOV-1999

```

```

DEFINITION Homo sapiens clone RP11-2P11, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011262
VERSION AC011262.2 GI:6454026
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145346)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-2P11
TITLE Unpublished
REFERENCE 2 (bases 1 to 145346)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 19, 1999 this sequence version replaced gi:6012128.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L2689
Center clone name: 2_P11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 97751 bases at least Q40
Consensus quality: 121236 bases at least Q30
Consensus quality: 136150 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 145346; sum-of-contigs
Quality coverage: 3.1 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1091: contig of 1091 bp in length
* gap of unknown length
* 1092 2456: contig of 1365 bp in length
* gap of unknown length
* 2457 3980: contig of 1524 bp in length
* gap of unknown length
* 3981 6275: contig of 2295 bp in length
* gap of unknown length
* 6276 8952: contig of 2677 bp in length
* gap of unknown length

```

```

* 8953 11694: contig of 2742 bp in length
* gap of unknown length
* 11695 12820: contig of 1126 bp in length
* gap of unknown length
* 12821 15352: contig of 2532 bp in length
* gap of unknown length
* 15353 19536: contig of 4184 bp in length
* gap of unknown length
* 19537 23248: contig of 3712 bp in length
* gap of unknown length
* 23249 28780: contig of 5532 bp in length
* gap of unknown length
* 28781 32808: contig of 4028 bp in length
* gap of unknown length
* 32809 37816: contig of 5008 bp in length
* gap of unknown length
* 37817 43068: contig of 5252 bp in length
* gap of unknown length
* 43069 48045: contig of 4977 bp in length
* gap of unknown length
* 48046 56867: contig of 8822 bp in length
* gap of unknown length
* 56868 67910: contig of 11043 bp in length
* gap of unknown length
* 67911 81484: contig of 13574 bp in length
* gap of unknown length
* 81485 99620: contig of 18136 bp in length
* gap of unknown length
* 99621 117196: contig of 17576 bp in length
* gap of unknown length
* 117197 145346: contig of 28150 bp in length.
* Location/Qualifiers
* 1. 145346
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /clone="RP11-2P11"
* /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 47197 a 25901 c 27079 g 45152 t 17 others
ORIGIN

Query Match 1.7% Score 45; DB 43; Length 145346;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Db 99962 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 100006

RESULT 44
AC015500/c
LOCUS AC015500 156730 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone RP11-21G19, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC015500
VERSION AC015500.2 GI:6553985
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156730)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-21G19
TITLE Unpublished
REFERENCE 2 (bases 1 to 156730)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```



```
* 28595 31448: contig of 2854 bp in length
*      gap of unknown length
* 31449 35732: contig of 4284 bp in length
*      gap of unknown length
* 35733 40093: contig of 4361 bp in length
*      gap of unknown length
* 40094 44260: contig of 4167 bp in length
*      gap of unknown length
* 44261 49519: contig of 5259 bp in length
*      gap of unknown length
* 49520 55034: contig of 5515 bp in length
*      gap of unknown length
* 55035 63390: contig of 8356 bp in length
*      gap of unknown length
* 63391 70014: contig of 6624 bp in length
*      gap of unknown length
* 70015 75415: contig of 5401 bp in length
*      gap of unknown length
* 75416 82425: contig of 7010 bp in length
*      gap of unknown length
* 82426 90586: contig of 8161 bp in length
*      gap of unknown length
* 90587 97723: contig of 7137 bp in length
*      gap of unknown length
* 97724 106002: contig of 8279 bp in length
*      gap of unknown length
* 106003 115333: contig of 9331 bp in length
*      gap of unknown length
* 115334 125768: contig of 10435 bp in length
*      gap of unknown length
* 125769 140335: contig of 14567 bp in length
*      gap of unknown length
* 140336 155326: contig of 14991 bp in length
*      gap of unknown length
* 155327 162504: contig of 7178 bp in length.
```

FEATURES

```
source
1..162504
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11"
  /clone="397_P_10"
  /clone_lib="RFCl-11 Human Male BAC"
BASE COUNT 51669 a 29955 c 29241 g 50694 t 945 others
ORIGIN
```

```
Query Match 1.7% Score 45; DB 42; Length 162504;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAA 2688
      |.....|
Db 39562 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAA 39518
```

Search completed: April 6, 2000, 14:55:06
Job time: 27491 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:17:39 ; Search time 6421.81 Seconds
(without alignments)
1338.160 Million cell updates/sec

Title: US-09-090-672B-6
Perfect score: 2276
Sequence: 1 CTGAACTGGGAGTCAGGTGG.....AAAAAAAAAAAAAAAAAAAA 2276

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST.*
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: gb_est1.*
21: gb_est2.*
22: gb_est3.*
23: gb_est4.*
24: gb_est5.*
25: gb_est6.*
26: gb_est7.*
27: gb_est8.*
28: gb_est9.*
29: gb_est10.*
30: gb_est11.*
31: gb_est12.*
32: gb_est13.*
33: gb_est14.*
34: gb_est15.*
35: gb_est16.*
36: gb_est17.*
37: gb_est18.*
38: gb_est19.*
39: gb_est20.*
40: gb_est21.*
41: gb_est22.*
42: gb_est23.*
43: gb_est24.*

44: gb_est25.*
45: gb_est26.*
46: gb_est27.*
47: gb_est28.*
48: gb_est29.*
49: gb_est30.*
50: gb_est31.*
51: gb_est32.*
52: em_est20.*
53: em_est21.*
54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*
59: gb_est33.*
60: gb_est34.*
61: gb_est35.*
62: gb_est36.*
63: gb_est37.*
64: gb_est38.*
65: em_est27.*
66: em_est28.*
67: em_est29.*
68: em_est30.*
69: gb_est39.*
70: gb_est40.*
71: gb_est41.*
72: gb_est42.*
73: gb_est43.*
74: gb_est44.*
75: em_est31.*
76: em_est32.*
77: em_est33.*
78: em_est34.*
79: gb_gss1.*
80: gb_gss2.*
81: gb_gss3.*
82: gb_gss4.*
83: em_gss1.*
84: em_gss2.*
85: em_gss3.*
86: em_gss4.*
87: gb_gss5.*
88: gb_gss6.*
89: gb_gss7.*
90: gb_gss8.*
91: gb_gss9.*
92: em_gss5.*
93: em_gss6.*
94: em_gss7.*
95: em_gss8.*
96: em_gss9.*
97: em_gss10.*
98: em_gss11.*
99: gb_gss10.*
100: gb_gss11.*
101: em_gss12.*
102: gb_gss12.*
103: gb_gss13.*
104: gb_gss14.*
105: gb_gss15.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%	No.	Score	Match	Length	DB	ID	Description
c	1	453	19.9	464	49	Ar660113			Ar660113 we56g02.x

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137760.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
FEATURES
Location/Qualifiers
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2345468"
/clone_lib="Soares_thymus_NHFTh"
/dev_stage="fetal"
/lab_host="DH10B (phage-resistant)"
/note="Organ: thymus, pooled; Vector: p7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCGAAGTCGGGCGCGCAAGCTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 173 a 93 c 59 g 153 t
ORIGIN

Query Match 19.8%; Score 450; DB 50; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.4e-185;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TTGTGTATGGTGTGGTGAACCTGGTTTAAATTAGTGAACCTGCTGAGACAGACGCTAT 1869
DB 450 TTGTGTATGGTGTGGTGAACCTGGTTTAAATTAGTGAACCTGCTGAGACAGACGCTAT 391
QY 1870 TCTCATGTACTGGCAGACCTGATTTCTGAGCATTTAATATGATGCCCTGGGAGTACA 1929
DB 390 TCTCATGTACTGGCAGACCTGATTTCTGAGCATTTAATATGATGCCCTGGGAGTACA 331
QY 1930 AAAGTGGAGTGGCCTGAGTAATGCAATATGGTGGTTTACCATTCTTTGAGGTAAAG 1989
DB 330 AAAGTGGAGTGGCCTGAGTAATGCAATATGGTGGTTTACCATTCTTTGAGGTAAAG 271
QY 1990 CATCATGAACCTTGTAAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTGCATAGGTT 2049
DB 270 CATCATGAACCTTGTAAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTGCATAGGTT 211
QY 2050 TAAATAATTTTAAATATATGCTTCGAGTTTAAATTTGTAATAGCGTAACATAATTTAACT 2109
DB 210 TAAATAATTTTAAATATATGCTTCGAGTTTAAATTTGTAATAGCGTAACATAATTTAACT 151
QY 2110 CTATAATGTGTTTCATCTCGGAATATCTTAACATATGAATTAATGTTTGCATGTTCACTT 2169
DB 150 CTATAATGTGTTTCATCTCGGAATATCTTAACATATGAATTAATGTTTGCATGTTCACTT 91
QY 2170 CAAGAGCCTTTTTTTGAAAAAGCTTTTTTTTGAATCATCAAGCTTTCACATTTAAAT 2229
DB 90 CAAGAGCCTTTTTTTGAAAAAGCTTTTTTTTGAATCATCAAGCTTTCACATTTAAAT 31
QY 2230 AAAGTGTGTTGAAGCTTTATTTAAAAAAA 2259
DB 30 AAAGTGTGTTGAAGCTTTATTTAAAAAAA 1

RESULT 3
AW183584/c

LOCUS AW183584 494 bp mRNA EST 18-NOV-1999
DEFINITION XJ77g10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663298 3', mRNA sequence.
ACCESSION AW183584
VERSION AW183584.1 GI:6452098
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 494)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135156.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
FEATURES
Location/Qualifiers
1..494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663298"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCL-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 176 a 100 c 64 g 154 t
ORIGIN

Query Match 19.5%; Score 443; DB 74; Length 494;
Best Local Similarity 99.8%; Pred. No. 6.8e-182;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1755 TGGGCCATATGAAAGGCTAAGCTTCACTGTAAATAATAACTGGGAATTCCTGGTTGTG 1814
DB 494 TGGGCCATATGAAAGGCTAAGCTTCACTGTAAATAATAACTGGGAATTCCTGGATTGTG 435
QY 1815 TATGGCTGCTGACCTGGTTTAAATAGTGAACCTGCTGAGACAGACAGCTATTCTCC 1874
DB 434 TATGGCTGCTGACCTGGTTTAAATAGTGAACCTGCTGAGACAGACAGCTATTCTCC 375
QY 1875 ATGTAAGTGGCAGAGCTGATTTCTGAGCATTTAATATGATGCCGTGGAGTACAAAGT 1934
DB 374 ATGTAAGTGGCAGAGCTGATTTCTGAGCATTTAATATGATGCCGTGGAGTACAAAGT 315
QY 1935 GAGTGTGGCCTGAGTAATGCAATATGGGTGTTTACATTTCTTAGGTAAGAAAGCATCA 1994
DB 314 GAGTGTGGCCTGAGTAATGCAATATGGGTGTTTACATTTCTTAGGTAAGAAAGCATCA 255
QY 1995 CATGAACCTGTAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTGCATAGGTTTAA 2054
DB 254 CATGAACCTGTAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTGCATAGGTTTAA 195
QY 2055 ATTTTAAATATATGGCTTGAGTTTAAATGTAAATAGCGGTAACATAATTTTAACTCTATA 2114
DB 194 ATTTTAAATATATGGCTTGAGTTTAAATGTAAATAGCGGTAACATAATTTTAACTCTATA 135

```
QY 2115 ATGTGTTTCATTCGGAATAAATCCTAAACATATGAATTAATGTTTGCATGTTCCACTTCCAAG 2174
|||||
Db 134 ATGTGTTTCATTCGGAATAAATCCTAAACATATGAATTAATGTTTGCATGTTCCACTTCCAAG 75
|||||
QY 2175 AGCCCTTTTTCAGAAAAAGCTTTTTCATCATCAAGTCTTTCACACATTTAAATAAAGT 2234
|||||
Db 74 AGCCCTTTTTCAGAAAAAGCTTTTTCATCATCAAGTCTTTCACACATTTAAATAAAGT 15
|||||
QY 2235 GTTTGAAAGCTTTA 2248
|||||
Db 14 GTTTGAAAGCTTTA 1
|||||

RESULT 4
AI338791/c 493 bp mRNA EST 13-FEB-1999
LOCUS qt53f04.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1951711 3', mRNA sequence.
ACCESSION AI338791
VERSION AI338791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 244 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 469.
Location/Qualifiers
1. 493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1951711"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."
BASE COUNT 175 a 97 c 64 g 157 t
ORIGIN

Query Match 19.3%; Score 439; DB 45; Length 493;
Best Local Similarity 100.0%; Pred. No. 3.7e-180;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TTGTGTATGGGTGGTGAACCTGGTGTATTAATAGTGAACCTGCGAGACAGAGCTAT 1869
|||||
Db 442 TTGTGTATGGGTGGTGAACCTGGTGTATTAATAGTGAACCTGCGAGACAGAGCTAT 383
|||||
QY 1870 TCTCCATGTAACGGCAAGACCTGATTTCTGACATTTAATATGGATGCGGTGGAGATACA 1929
|||||
```

```
Db 382 TCTCCATGTAACGGCAAGACCTGATTTCTGAGCATTTAATATGGATGCGGTGGAGATACA 323
QY 1930 AAAGTGGAGTGGCCCTGAGTAATGATGATGATGGTGGTTTACCATTCTTGGAGGTAAGAAG 1989
|||||
Db 322 AAAGTGGAGTGGCCCTGAGTAATGATGATGATGGTGGTTTACCATTCTTGGAGGTAAGAAG 263
|||||
QY 1990 CATCACATGAACCTGTAAAGGAATTTAAAAATCCCTACTTTCATATAAGTTCGATAGGTT 2049
|||||
Db 262 CATCACATGAACCTGTAAAGGAATTTAAAAATCCCTACTTTCATATAAGTTCGATAGGTT 203
|||||
QY 2050 TAATAATTTTAAATATATGAGTGGTGGTAAATTTGTAATAGCGGTAACTAATTTAACT 2109
|||||
Db 202 TAATAATTTTAAATATATGAGTGGTGGTAAATTTGTAATAGCGGTAACTAATTTAACT 143
|||||
QY 2110 CTATAATGTTTCATCTCGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTT 2169
|||||
Db 142 CTATAATGTTTCATCTCGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTT 83
|||||
QY 2170 CCAAGAGCCCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAAAT 2229
|||||
Db 82 CCAAGAGCCCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAAAT 23
|||||
QY 2230 AAAGTGTGTAAGCTTTA 2248
|||||
Db 22 AAAGTGTGTAAGCTTTA 4
|||||

RESULT 5
AI399641/c 697 bp mRNA EST 30-MAR-1999
LOCUS th28a09.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2119576 3',
DEFINITION similar to TR:Q13545 Q13545 SIGNALING INOSITOL POLYPHOSPHATE 5
PHOSPHATASE SIP-145 ;, mRNA sequence.
ACCESSION AI399641
VERSION AI399641
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2979819.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 832 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 339.
Location/Qualifiers
1. 697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2119576"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: p7T3D-Pac (Pharmacia)
with a modified polylinker; plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
```


circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (c20615). 985608-986759, 1101192-1101959, and 1217928-1220615. Subtraction by Bento Soares and M. Fatima Bonaído. a 1149 c 116 g 236 t 2 others

BASE COUNT	ORIGIN
194 a	149 c
116 g	236 t
2 others	

```
Query Match      19.0%; Score 432; DB 46; Length 697;
Best Local Similarity 99.7%; Pred. No. 3.5e-177;
Matches 602; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

QY	217	GGCGTGACTGCTATGTGTGCTGTATCACCGGTTACATTTTATACATACCGAGTGCTCCAG	276
Db	604	GGCGTGACTGCTATGTGTGCTGTATCACCGGTTACATTTTATACATACCGAGTGCTCCAG	545
QY	277	ACAGAAACAGGTTCTTGGAGTCTGAGACACACCTGGGTACATAAAAGATATTTCCGG	336
Db	544	ACAGAAACAGGTTCTTGGAGTCTGAGACACACCTGGGTACATAAAAGATATTTCCGG	485
QY	337	AAATAAAAATCTCATTTACGATTTTCAGAGCCAGATCAAGGCATTTGTAATACCTCTG	396
Db	484	AAATAAAAATCTCATTTACGATTTTCAGAGCCAGATCAAGGCATTTGTAATACCTCTG	425
QY	397	CAGTATCCAGTTGAGAAAGTCCCTCAGCTAGAAGTACACAGGTACTACAGGGATAAGA	456
Db	424	CAGTATCCAGTTGAGAAAGTCCCTCAGCTAGAAGTACACAGGTACTACAGGGATAAGA	365
QY	457	GAAGATCCTGATGCTGCTGAAAGCCCCATGAAGAAAAATAAAACACCTTGTCATTTAT	516
Db	364	GAAGATCCTGATGCTGCTGAAAGCCCCATGAAGAAAAATAAAACACCTTGTCATTTAT	305
QY	517	TTTCTATAATTAATAATATGCTTAAGTCTTATATATTGTAGATAATACAGTTCGGTGAGC	576
Db	304	TTTCTATAATTAATAATATGCTTAAGTCTTATATATTGTAGATAATACAGTTCGGTGAGC	245
QY	577	TACAAATGCATTTCTAAAGCCATTCGTAGTCTCTGTAATGGAAGCATCTAGCATGCTGCAA	636
Db	244	TACAAATGCATTTCTAAAGCCATTCGTAGTCTCTGTAATGGAAGCATCTAGCATGCTGCAA	185
QY	637	AGCTGAAATGGACTTTTGTACATAGTGAGGAGCTTTGAAACGAGGATTGGG-AAAAGTAA	695
Db	184	AGCTGAAATGGACTTTTGTACATAGTGAGGAGCTTTGAAACGAGGATTGGGAAAAAGTAA	125
QY	696	TTCCGTAGGTTATTTTCAGTTATTATTTACAAATGGGAAACAAAAGGATAATGAATAC	755
Db	124	TTCCGTAGGTTATTTTCAGTTATTATTTACAAATGGGAAACAAAAGGATAATGAATAC	65
QY	756	TTTATAAGGATTAAATGCTCAATTTCTGCCAATAATAATAAAAATATCTCAGTTTTTG	815
Db	64	TTTATAAGGATTAAATGCTCAATTTCTGCCAATAATAATAAAAATATCTCAGTTTTTG	5
QY	816	TGAA	819
Db	4	TGAA	1

RESULT	6
AI377357/c	
LOCUS	
DEFINITION	AI377357 447 bp mRNA EST te56f01.xl Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2091577 3' similar to TR:Q13545 Q13545 SIGNALING INOSITOL POLYPHOSPHATE 5 PHOSPHATASE SIP-145 ; , mRNA sequence.
ACCESSION	AI377357
VERSION	AI377357.1 GI:4187210
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 447)

```

AA243694/c
LOCUS      AA243694      421 bp      mRNA      EST      07-MAR-1997
DEFINITION zif6808.s1 Soares.NhMMPu_S1 Homo sapiens cDNA clone IMAGE:668582
            3', mRNA sequence.
ACCESSION  AA243694
VERSION    AA243694.1 GI:1874486
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 421)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevasaki,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT    On May 9, 1995 this sequence version replaced gi:802693.
            Contact: Wilton RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. Et from Amersham
            High quality sequence stop: 373.
FEATURES   Location/Qualifiers
            source          1..421
                        /organism="Homo sapiens"
                        /db_xref="GB:5562549"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:668582"
                        /clone_lib="Soares.NhMMPu_S1"
                        /tissue_type="Pooled human melanocyte, fetal heart, and
                        pregnant uterus"
                        /lab_host="DH10B"
                        /note="Organ: mixed (see below); Vector: pT73D-Pac
                        (Pharmacia) with a modified polylinker; Site_1: Not 1;
                        Site_2: Eco RI; Equal amounts of plasmid DNA from three
                        normalized libraries (melanocyte 2NbHM, pregnant uterus
                        NhMMPu, and fetal heart NBHL19W) were mixed, and ss circles
                        were made in vitro. Following HAP purification, this DNA
                        was used as tracer in a subtractive hybridization
                        reaction. The driver was PCR-amplified cDNAs from pools of
                        5,000 clones made from the same 3 libraries. The pools
                        consisted of I.M.A.G.E. clones 260232-265223,
                        340488-345479, and 484488-489479."
BASE COUNT 152 a 79 c 55 g 135 t
ORIGIN
            16.7%; Score 381; DB 30; Length 421;
            Best Local Similarity 100.0%; Pred. No. 5.7e-155;
            Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1868 ATTCTCATGTCGACGACCTGATTTCTGAGCATTTAATGATGCGCTGGAGTA 1927
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 383 ATTTCCATGTCGACGACCTGATTTCTGAGCATTTAATGATGCGCTGGAGTA 324
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1928 CAAAAGTGGAGTGGCTGAGTAATGATGATGGTGGTTTACCATTCTTTGAGTAA 1987
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 323 CAAAAGTGGAGTGGCTGAGTAATGATGATGGTGGTTTACCATTCTTTGAGTAA 264
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1988 AGCATCATGAACTTCTAAGGAATTTAAATCCTACTTTTCATATGAATGATGATG 2047
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 263 AGCATCATGAACTTCTAAGGAATTTAAATCCTACTTTTCATATGAATGATGATG 204
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 2048 TTTAATAATTTTAAATATGCGTGGTTAAATTTAAATAGGCGTAACATAATTTAA 2107
            |||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db 203 TTTAATAATTTTAAATATATGCGTTGAGTTTAAATTTGTAATAGCGGTAACATAATTTAA 144
Qy 2108 CTCTATAATGTTTCATCTCGGAATAATCCTAAACATATGAATTTATGATGTTGCATGTTAC 2167
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 143 CTCTATAATGTTTCATCTCGGAATAATCCTAAACATATGAATTTATGATGTTGCATGTTAC 84
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 2168 TTCCAGAGCGCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAA 2227
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 83 TTCCAGAGCGCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAA 24
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 2228 ATAAGTGTGTTGAAGCTTTA 2248
            |||||||||||||||||||||||||||
Db 23 ATAAGTGTGTTGAAGCTTTA 3
            |||||||||||||||||||||||

RESULT 8
AI215447/c
LOCUS      AI215447      360 bp      mRNA      EST      22-OCT-1998
DEFINITION qh11e04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1844382 3', mRNA sequence.
ACCESSION  AI215447
VERSION    AI215447
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 360)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Jan 19, 1998 this sequence version replaced gi:2043866.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -400P from Gibco
            High quality sequence stop: 347.
FEATURES   Location/Qualifiers
            source          1..360
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1844382"
                        /clone_lib="Soares_NFL_T_GBC_S1"
                        /lab_host="DH10B"
                        /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
                        a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
                        Equal amounts of plasmid DNA from three normalized
                        libraries (fetal lung NBHL19W, testis NHT, and B-cell
                        NCI-CGAP GCB1) were mixed, and ss circles were made in
                        vitro. Following HAP purification, this DNA was used as
                        tracer in a subtractive hybridization reaction. The driver
                        was PCR-amplified cDNAs from pools of 5,000 clones made
                        from the same 3 libraries. The pools consisted of
                        I.M.A.G.E. clones 297480-302087, 682632-687239,
                        726408-728711, and 729096-731399. Subtraction by Bento
                        Soares and M. Fatima Bonaldo."
BASE COUNT 135 a 64 c 45 g 116 t
ORIGIN
            15.8%; Score 359; DB 43; Length 360;
            Best Local Similarity 100.0%; Pred. No. 2.1e-145;
            Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1890 CTGATTTCTGAGCATTTAATGATGCGCTGGAGTACAAAGTGGAGTGGCCTGAG 1949
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 360 CTGATTTCTGAGCATTTAATGATGCGCTGGAGTACAAAGTGGAGTGGCCTGAG 301
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1950 TAATGCAATGATGGTGGTTTACCATTTCTTGTAGGTTAAAGCATCATCAATGTAAG 2009
            |||||||||||||||||||||||||||||||||||||||||||||||||||

```

QY	1026	TCTGAAAGCAACCTTCTTCGTGCCTAGTGTTCTGTATATTGGACAGTAGTAATAATCCACAGACCAAC	1085
Db	475	TCTGAAAGCAACCTTCTTCGTGCCTAGTGTTCTGTATATTGGACAGTAGTAATAATCCACAGACCAAC	416
QY	1086	CTGGAGCTTGAAAATCTTATAATTTAAAAATATGCTCTAAACAATGTTTATCGTATTGTATGC	1145
Db	415	CTGGAGCTTGAAAATCTTATAATTTAAAAATATGCTCTAAACAATGTTTATCGTATTGTATGC	356
QY	1146	TACAGAGTTGAAATGTATATACAAATCCAATGAATGAAATGAGTTTTCTTTTCATTACTCCTC	1205
Db	355	TACAGAGTTGAAATGTATATACAAATCCAATGAATGAAATGAGTTTTCTTTTCATTACTCCTC	296
QY	1206	TGCCCCAGCTGTTTCTTACTACATGGAAGACCTCATTTTGAAGGGAAAAATTTTCAGCAGCTGC	1265
Db	295	TGCCCCAGCTGTTTCTTACTACATGGAAGACCTCATTTTGAAGGGAAAAATTTTCAGCAGCTGC	236
QY	1266	AGCTCATGAGTAACTGATTTGTAAACAAGCCTCTCTTTTAAAGTACCCTACAAAACCACTG	1325
Db	235	AGCTCATGAGTAACTGATTTGTAAACAAGCCTCTCTTTTAAAGTACCCTACAAAACCACTG	176
QY	1326	GAAAGTTTATGGTTGATTTATTTT 1350	
Db	175	GAAAGTTTATGGTTGATTTATTTT 151	
RESULT 10			
N89899/c			
LOCUS			
DEFINITION			
N89899 437 bp mRNA EST 02-APR-1996			
zb22c09.s1 Soares fetal_lung_MbHL19W Homo sapiens cDNA clone			
IMAGE:302800 3' similar to SW:ABL_DROME P00522 TYROSINE-PROTEIN-KINASE DASH/ABL ; mRNA sequence.			
ACCESSION N89899			
VERSION N89899.1 Gi:1443226			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
AUTHORS Hillier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.			
TITLE The WashU-Merck EST Project			
JOURNAL Unpublished [1995]			
COMMENT On Apr 14, 1993 this sequence version replaced gi:693457. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: ET primer High quality sequence stop: 359. Location/Qualifiers 1. 437 /organism="Homo sapiens" /db_xref="GDB:1247732" /db_xref="taxon:9606" /clone="IMAGE:302800" /clone_lib="Soares_fetal_lung_MbHL19W" /dev_stage="19 weeks" /lab_host="dH10B (ampicillin resistant)" /note="organ: lung; Vector: p7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Is strand cDNA was primed with a Not I - oligo(dT) primer 15'-TGTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTTTTTT double-stranded cDNA was size selected, ligated to Eco adapters (Pharmacia), digested with Not I and cloned into plasmids (Pharmacia).			
FEATURES			
source			

the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBH19W."

BASE COUNT 103 a 108 c 93 g 132 t 1 others
ORIGIN

Query Match 13.8%; Score 314; DB 25; Length 437;
Best Local Similarity 99.8%; Pred. No. 6.4e-126;
Matches 434; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 95 ACCAGGCATGGACGAGTGGCTGTATCATGGCAAAATCAGCAGGAAACCGCGGAGA 154
|||||
Db 435 ACCAGGCATGGACGAGTGGCTGTATCATGGCAAAATCAGCAGGAAACCGCGGAGA 376
|||||

QY 155 AGCTCCTGCTGGCAGTGGCTGGATGGCAGCTATTGTGAGGGACAGCAGAGCGTGC 214
|||||
Db 375 AGCTCCTGCTGGCAGTGGCTGGATGGCAGCTATTGTGAGGGACAGCAGAGCGTGC 316
|||||

QY 215 CAGCGGTGACTGCTATGCTGTATCAGCGTTACATTTATACATACCAGTGTCC 274
|||||
Db 315 CAGCGGTGACTGCTATGCTGTATCAGCGTTACATTTATACATACCAGTGTCC 256
|||||

QY 275 AGACAGAAACAGGTTCTTGAGAGTGGCTGAGACAGCACCTGGGGTACATAAAGATATTCC 334
|||||
Db 255 AGACAGAAACAGGTTCTTGAGAGTGGCTGAGACAGCACCTGGGGTACATAAAGATATTCC 196
|||||

QY 335 GGAATAAATAATCTCATTTTC-AGCATTTGAGAACGAGATCAGGCAATTTGATACCT 393
|||||
Db 195 GGAATAAATAATCTCATTTTCGAGCATTTTTCAGAACGAGATCAGGCAATTTGATACCT 136
|||||

QY 394 CTGAGTATCCAGTTGAGAGAGTCTCTAGCTAGAGTACACAGTACTACAGGGATA 453
|||||
Db 135 CTGAGTATCCAGTTGAGAGAGTCTCTAGCTAGAGTACACAGTACTACAGGGATA 76
|||||

QY 454 AGAGAATCCTGATGCTGCTGAAAGCCCATGAAGAAAAATAAACACCTTGTACTT 513
|||||
Db 75 AGAGAATCCTGATGCTGCTGAAAGCCCATGAAGAAAAATAAACACCTTGTACTT 16
|||||

QY 514 TATTTCTATAATT 528
|||||
Db 15 TATTTCTATAATT 1

RESULT 11
AI203763/C 243 bp mRNA EST 27-JAN-1999
LOCUS qf76e08.x1 Soares.fetal_lung_NBHL19W Homo sapiens CDNA clone
DEFINITION IMAGE:1755974 3', mRNA sequence.
ACCESSION AI203763
VERSION AI203763.1 GI:3756369
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 243)
AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152111.
Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1066 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 225.

Location/Qualifiers
1. .243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1755974"
/clone.lib="Soares.fetal_lung_NBHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBH19W."

BASE COUNT 88 a 38 c 43 g 74 t
ORIGIN

Query Match 10.6%; Score 242; DB 43; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.4e-94;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1109 TAAATATGCTCTAAACATGTTTATCGTATTGATGCTACAGGATTTGAAATTTCTATTAC 1168
Db 243 TAAATATGCTCTAAACATGTTTATCGTATTGATGCTACAGGATTTGAAATTTCTATTAC 184
QY 1169 AAATCCAATGAATGAGTGTCTTTTTCATTTACCTCTGCCCCAGTGTCTTCTACTACAT 1228
Db 183 AAATCCAATGAATGAGTGTCTTTTTCATTTACCTCTGCCCCAGTGTCTTCTACTACAT 124
QY 1229 GGAAGACCTCATTTTGAAGGAAATTTTCAGAGCTGAGCTCATGAGTAACGTATTGTA 1288
Db 123 GGAAGACCTCATTTTGAAGGAAATTTTCAGAGCTGAGCTCATGAGTAACGTATTGTA 64
QY 1289 ACAAGCTCCTTTTAAAGTAACCCCTACAAAACACCTGAAAAGTTTATGTTGTTATTATT 1348
Db 63 ACAAGCTCCTTTTAAAGTAACCCCTACAAAACACCTGAAAAGTTTATGTTGTTATTATT 4
QY 1349 TT 1350
Db 3 TT 2

RESULT 12
AA233116
LOCUS 208 bp mRNA EST 28-FEB-1997
DEFINITION Zr68e08.r1 Soares_NHMPu_S1 Homo sapiens CDNA clone IMAGE:668582
5', mRNA sequence.
ACCESSION AA233116
VERSION AA233116.1 GI:1856110
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 208)
AUTHORS Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Dec 3, 1996 this sequence version replaced gi:1122614.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m13 rev2 Et from Amersham

High quality sequence stop: 192.

FEATURES

source

Location/Qualifiers

1..208

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:668582"

/clone_lib="Soares_NHRMPu_S1"

/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"

/lab_host="DH10B"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NDHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones, 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 73 a 35 c 38 g 62 t

ORIGIN

Query Match 7.1%; Score 162; DB 30; Length 208;
Best Local Similarity 100.0%; Pred. No. 7.4e-60;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 CTGATGTCGCTGAAGCCCATGAGAAATAAACAACCTGTACTTTATTTCTAT 523
|||||
DB 47 CTGATGTCGCTGAAGCCCATGAGAAATAAACAACCTGTACTTTATTTCTAT 106
|||||

QY 524 AATTAAATATATGCTAAGTCTTATATATGTTAGATATACAGTTCGGTGAGCTACAAAT 583
|||||

DB 107 AATTAAATATATGCTAAGTCTTATATATGTTAGATATACAGTTCGGTGAGCTACAAAT 166
|||||

QY 584 GCATTCTAAGCCATTGTAGTCTGTATGGAAGCATCTAG 625
|||||

DB 167 GCATTCTAAGCCATTGTAGTCTGTATGGAAGCATCTAG 208
|||||

RESULT 13

AA190204/c

LOCUS

TH075 HTCDD1 Homo sapiens CDNA 5'/3' similar to Unknown, mRNA

sequence.

ACCESSION

AA190204

VERSION

AA190204.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 166)

Sohn, U., Park, D.S., Lee, C.M., Cho, W.K., Ahn, H.J., Lee, M.Y.,

Hwang, M.Y., and Jin, S.W.

Human HTCDD1 library cDNAs

Unpublished (1994)

TITLE

Human HTCDD1 library cDNAs

JOURNAL

On Sep 12, 1996 this sequence version replaced gi:1407222.

COMMENT

Contact: Uik

Sohn, D.S. Park, C.M. Lee, W.K. Cho, H.J. Ahn, M.Y. Lee, M.Y. Hwang, S.W. Jin

Laboratory of Molecular Biology

Kyungpook National University

Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701, Korea

Tel: 82-053-950-5382

Fax: 82-053-955-5327

Email: usohn@bh.kyungpook.ac.kr

Seq primer: M13 Reverse/SK primer.

FEATURES

source

Location/Qualifiers

1..166

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HTCDD1"

/lab_host="XLI-Blue"

/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2: EcoRI; Poly(A)-mRNA from the 2-year old male fetal thymus, oligo(dT) priming, EcoRI cloning in the vector pBluescript (Stratagene)."

BASE COUNT 66 a 18 c 24 g 58 t

ORIGIN

Query Match 6.8%; Score 155; DB 29; Length 166;
Best Local Similarity 100.0%; Pred. No. 8.8e-57;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2098 CTAATTTAACTCTATAATGTTTCATTCCTGGAATAATCCTAAACATATGAATATGTTT 2157
|||||

DB 166 CTAATTTAACTCTATAATGTTTCATTCCTGGAATAATCCTAAACATATGAATATGTTT 107
|||||

QY 2158 GCATGTCACATCCAGAGCCCTTTTGAAGGCTTTTGAATCATCAAGCTTT 2217
|||||

DB 106 GCATGTCACATCCAGAGCCCTTTTGAAGGCTTTTGAATCATCAAGCTTT 47
|||||

QY 2218 TCACATTTAAATAAGTGTGTTGAAAGCTTTTATTTA 2252
|||||

DB 46 TCACATTTAAATAAGTGTGTTGAAAGCTTTTATTTA 12
|||||

RESULT 14

AW001042/c

LOCUS

wr91h07.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2495101 3',

mrna sequence.

ACCESSION

AW001042

VERSION

AW001042.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 359)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

JOURNAL

On May 9, 1996 this sequence version replaced gi:1132712.

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from Gibco.

Location/Qualifiers

1..359

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2495101"

/clone_lib="NCI-CGAP_Kid11"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 132376-132391, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldi.

BASE COUNT 120 a 60 c 60 g 119 t

ORIGIN

Query Match 5.4%; Score 122; DB 63; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.3e-42; Mismatches 0; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 570 GGTGAGCTACAATGCAATCTTAAGCCATGTAGTCTGTAAATGGAAGCATCTAGCATG 629
|||||
Db 246 GGTGAGCTACAATGCAATCTTAAGCCATGTAGTCTGTAAATGGAAGCATCTAGCATG 187
|||||

Qy 630 TCGTCAAGCTGAATGGACCTTTTGTACATAGTGAGGAGCTTTGAAACGAGGATTGGAA 689
|||||

Db 186 TCGTCAAGCTGAATGGACCTTTTGTACATAGTGAGGAGCTTTGAAACGAGGATTGGAA 127
|||||

Qy 690 AA 691
||

Db 126 AA 125

RESULT 15
AA585096
LOCUS 172 bp mRNA EST 09-SEP-1997
DEFINITION ATH247 HTCdll Homo sapiens CDNA 5'/3', mRNA sequence.
ACCESSION AA585096
VERSION AA585096.1 GI:2384984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172)
AUTHORS Sohn,U., Park,D.S., Lee,C.M., Cho,W.K., Ahn,H.J., Lee,M.Y., Hwang,M.Y. and Jin,S.W.
TITLE Human HTCdll library cDNAs
JOURNAL Unpublished (1994)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1294001.
Contact: Uik
Sohn,D.S.Park,C.M.Lee,W.K.Cho,H.J.Ahn,M.Y.Lee,M.Y.Hwang,S.W.Jin
Laboratory of Molecular Biology
Kyungpook National University
Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701,
Korea
Tel: 82-053-950-5382
Fax: 82-053-955-5327
Email: usohn@kh.kyungpook.ac.kr
Seq primer: M13 Reverse/SK primer.

FEATURES
source
1..172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HTCdll"
/lab_host="XLI-Blue"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: EcoRI; Poly(A)-mRNA from the 2-year old male fetal thymus, oligo(dT) priming, EcoRI cloning in the vector pBluescript (Stratagene)."
BASE COUNT 54 a 27 c 27 g 64 t

ORIGIN

Query Match 5.2%; Score 119; DB 35; Length 172;

Best Local Similarity 100.0%; Pred. No. 3.6e-41;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2099 TAATTTAACTCTAATATGTGTTCTGGAATAATCCTAAACATATGAATTATGTTG 2158
|||||
Db 29 TAATTTAACTCTAATATGTGTTCTGGAATAATCCTAAACATATGAATTATGTTG 88
|||||

Qy 2159 CAGTTCACCTCCAAAGAGCCTTTTGTGAAAAAGCTTTTGTGATCATCAAGTCTT 2217
|||||
Db 89 CATGTTCACTCCAAAGAGCCTTTTGTGAAAAAGCTTTTGTGATCATCAAGTCTT 147
|||||

RESULT 16
AA354319
LOCUS 331 bp mRNA EST 21-APR-1997
DEFINITION Jurkat T-cells V Homo sapiens CDNA 5' end, mRNA sequence.
ACCESSION AA354319
VERSION AA354319.1 GI:2006639
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 331)
AUTHORS Adams,M.D., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,F.J., Dimke,D., Feng,D.F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Müssner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 12140200
COMMENT On May 8, 1995 this sequence version replaced gi:801515.
Other_ESrs: THCL48776
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1..331
/organism="Homo sapiens"
/db_xref="Arcc (inhost):154561"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells V"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 84 a 82 c 89 g 73 t 3 others

ORIGIN

Query Match 4.9%; Score 111; DB 32; Length 331;
Best Local Similarity 100.0%; Pred. No. 8e-38;


```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 308)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,X.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
12140200
Other_ESTs: THCL48776
On Apr 14, 1993 this sequence version replaced gi:638038.
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers
1..308
/organism="Homo sapiens"
/db_xref="ATCC (inhost):157953"
/db_xref="taxon:9606"
/clone_lib="Jurkat T-cells VI"
/cell_type="T-lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 69 a 70 c 87 g 76 t 6 others
ORIGIN
Query Match 2.0%; Score 46; DB 32; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAACGGGAGTCAGGTGCTGACTTGCTCCGTCGAGTAGCA 46
|||||
Db 146 CTGAACGGGAGTCAGGTGCTGACTTGCTCCGTCGAGTAGCA 191
|||||

RESULT 20
LOCUS AI030232/c 590 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C0-It-a-04-0-UI s1 UI-R-C0 Rattus norvegicus cDNA clone
UI-R-C0-It-a-04-0-UI 3', mRNA sequence.
ACCESSION AI030232
VERSION AI030232.1 GI:3248058
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 590)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900524.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics
Seq primer: M13 Forward
Location/Qualifiers
1..590
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C0-It-a-04-0-UI"
/clone_lib="UI-R-C0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
library is a subtracted library derived from the UI-R-A1
and UI-R-El libraries. The UI-R-A1 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The UI-R-El
library consisted of a mixture of individually tagged
normalized libraries constructed from 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which
allows identification of the library of origin of a clone
within the mixture. The subtracted library (UI-R-C0) was
constructed as follows: PCR amplified cDNA inserts from a
pool of UI-R-A1 and UI-R-El clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the pooled UI-R-A1 and UI-R-El library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C0
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."

BASE COUNT 163 a 116 c 109 g 202 t
ORIGIN
Query Match 1.9%; Score 43; DB 41; Length 590;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 TAAAAATCTCTTTCAGCATTTTCAGAGCCAGATCAAGGCAT 383
|||||
Db 488 TAAAAATCTCTTTCAGCATTTTCAGAGCCAGATCAAGGCAT 446
|||||

RESULT 21
LOCUS AI131797
DEFINITION uc37e03.r1 Soares mouse mammary gland mRNMG Mus musculus cDNA clone
IMAGE:1400188 5' similar to TR:Q13545 SIGNALING INOSITOL

```



```

POLYPHOSPHATE 5 PHOSPHATASE SIP-145 ;, mRNA sequence.
ACCESSION A1131797 GI:3601813
VERSION A1131797.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 271)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1407241.
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:911904
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 89.
FEATURES
Location/Qualifiers
1..271
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:400188"
/clone_lib="Soares mouse mammary gland NbMNG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: p7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru KO, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 66 a 63 c 75 g 67 t
ORIGIN
Query Match 1.5%; Score 35; DB 42; Length 271;
Best Local Similarity 100.0%; Pred. No. 8e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 265 CGAGTGTCCACAGACAACAGGTTCTTGGAGTGC 299
|||||
Db 203 CGAGTGTCCACAGACAACAGGTTCTTGGAGTGC 237
|||||
RESULT 22
AA255258 294 bp mRNA EST 14-MAR-1997
LOCUS mz77h04.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
DEFINITION IMAGE:719479 5' similar to TR:G1255352 SH2-CONTAINING
INOSITOL PHOSPHATASE SHIP. ;, mRNA sequence.
ACCESSION AA255258
VERSION AA255258.1 GI:1889847
KEYWORDS EST.
POLYPHOSPHATE 5 PHOSPHATASE SIP-145 ;, mRNA sequence.
ACCESSION A1464772 GI:4318802
VERSION A1464772.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 397)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
JOURNAL On Apr 14, 1993 this sequence version replaced gi:638194.
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:444975
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 275.
FEATURES
Location/Qualifiers
1..294
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:719479"
/clone_lib="Soares mouse lymph node NbMLN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TGTTACCAATCTGAAGTGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 80 a 69 c 76 g 69 t
ORIGIN
Query Match 1.5%; Score 35; DB 30; Length 294;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 265 CGAGTGTCCACAGACAACAGGTTCTTGGAGTGC 299
|||||
Db 84 CGAGTGTCCACAGACAACAGGTTCTTGGAGTGC 118
|||||
RESULT 23
A1464772 397 bp mRNA EST 09-MAR-1999
LOCUS mz77h04.y1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
DEFINITION IMAGE:719479 5' similar to TR:Q61181 Q61181 INOSITOL
POLYPHOSPHATE-5-PHOSPHATASE, 145 KDA ;, mRNA sequence.
ACCESSION A1464772
VERSION A1464772.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 397)

```

AUTHORS

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

On May 7, 1998 this sequence version replaced gi:3121624.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

This read is a RESSEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)

Putative full length read

vector to vector length is 643

Seq primer: -40RP from Gibco

High quality sequence stop: 362.

Location/Qualifiers

1. .397

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:719479"

/clone_lib="Soares mouse lymph node NbMLN"

/sex="male"

/tissue_type="lymph node"

/dev_stage="4 weeks"

/lab_host="DH10B"

/notes="Organ: lymph node; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5, TGTACCAATCTGAAGTGGAGCGCGGCGGATCTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

113 a 87 c 95 g 102 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 1.5%; Score 35; DB 47; Length 397;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 CGAGTGTCACAGACAGAACAGGTTCTTGGAGTGC 299

|||||

Db 84 CGAGTGTCACAGACAGAACAGGTTCTTGGAGTGC 118

RESULT 24

AI522369/c

LOCUS fb22ell.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA

DEFINITION AI522369 210 bp mRNA EST 18-MAR-1999

sequence.

ACCESSION AI522369

VERSION AI522369.1 GI:4436504

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Cyprininae; Rasborinae; Danio.

1 (bases 1 to 210)

REFERENCE

AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,

Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

WashU zebrafish EST Project 1998

Unpublished (1998)

On Mar 10, 1998 this sequence version replaced gi:2949061.

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and

ResourcenzentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)

Seq primer: 7 ET from Amersham

High quality sequence stop: 187.

Location/Qualifiers

1. .210

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone_lib="Zebrafish WashU MPIMG EST"

/sex="mixed"

/tissue_type="26 somite embryos, adult livers, shield

stage embryos"

/lab_host="XLI-blue MRF"

/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTCTAGATCGAGCGCGCGCCCTTTTTTTTTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss) adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 72 a 28 c 25 g 85 t

ORIGIN

Query Match

Best Local Similarity 1.4%; Score 33; DB 47; Length 210;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2244 CTTTATTTAAAAAATAAAAAAAAAAAAAA 2276

|||||

Db 33 CTTTATTTAAAAAATAAAAAAAAAAAAAA 1

RESULT 25

AU074326

LOCUS

DEFINITION

AU074326 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium

discoideum cDNA clone SSK377, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

237 bp mRNA EST 24-JUN-1999

GI:5180747

Dictyostellium discoideum.

Dictyostellium discoideum

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Dictyosteliida; Dictyostelium.
1 (bases 1 to 237)
Urushihara, H.
Developmental cDNA in Dictyostelium discoideum (1999)
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3246998.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT : Dictyostelium discoideum cDNA project in Japan.
Location/Qualifiers

FEATURES
source

1..237
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSK377"
/clone.lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
123 a 14 c 14 g 86 t

BASE COUNT
ORIGIN

Query Match 1.4%; Score 33; DB 51; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.00062;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
|||||...|||||
DB 99 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 131
|||||...|||||

RESULT 26
AI522765/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI522765 301 bp mRNA EST 18-MAR-1999
fb52b03.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
sequence.
AI522765
AI522765.1 GI:4436900
EST.
zebrafish.
Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 301)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
On Mar 10, 1998 this sequence version replaced gi:2949468.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimatendatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: 77 ET from Amersham
High quality sequence stop: 300.
Location/Qualifiers

source

1..301
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 sonite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACTAGTTCTAGATCGAGCGCGCCCTTTTITTTTITTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 88 a 47 c 56 g 110 t
ORIGIN

Query Match 1.4%; Score 33; DB 47; Length 301;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
|||||...|||||
DB 37 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 5
|||||...|||||

RESULT 27
AI501287/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI501287 420 bp mRNA EST 21-APR-1999
fc09b07.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3' similar to
contains element TATL repetitive element ;, mRNA sequence.
AI501287
AI501287.1 GI:4610448
EST.
zebrafish.
Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 420)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
On Mar 20, 1998 this sequence version replaced gi:2980338.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and
 RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)

Possible reversed clone: polyt not found

Seq primer: T7 ET from Amersham

High quality sequence stop: 408.

FEATURES

source
 1. 420
 /organization="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish WashU MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; lstr
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'pgactagttctagatcgagcgccgctttttttttttt3'];
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 108 a 91 c 67 g 151 t 3 others
 ORIGIN

Query Match 1.4%; Score 33; DB 48; Length 420;
 Best Local Similarity 100.0%; Pred. No. 0.00049;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
 ||||||||||||||||||||||||||||||||||||||||
 Db 45 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 13

RESULT 28
 AI721672/c
 LOCUS
 DEFINITION fc30g11.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
 sequence.
 AI721672
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Danio rerio
 Eukaryota; Chordata; Craniata; Vertebrata; Actinopterygii;
 Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 Cyprinidae; Cyprininae; Rasboreinae; Danio.
 1 (bases 1 to 424)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marr, M.,
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 On Mar 10, 1998 this sequence version replaced gi:2947976.
 Other_ESTs: fc30g11.y1
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)

Seq primer: T7 ET from Amersham
 High quality sequence stop: 423.

FEATURES

source
 1. 424
 /organization="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish WashU MPIMG EST"
 /sex="mixed"
 /tissue_type="26 somite embryos, adult livers, shield
 stage embryos"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; lstr
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'pgactagttctagatcgagcgccgctttttttttttt3'];
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 106 a 94 c 92 g 132 t
 ORIGIN

Query Match 1.4%; Score 33; DB 51; Length 424;
 Best Local Similarity 100.0%; Pred. No. 0.00049;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
 ||||||||||||||||||||||||||||||||||||||||
 Db 44 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 12

RESULT 29
 AI833180/c
 LOCUS
 DEFINITION at65f03.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone
 IMAGE:2376893 3', similar to gb:xl6100 NADH-UBIQUINONE
 OXIDOREDUCTASE 75 KD SUBUNIT PRECURSOR (HUMAN);, mRNA sequence.
 AI833180
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 553)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Lennon, G., Marr, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, E.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)

COMMENT

On Dec 20, 1995 this sequence version replaced gi:1134468.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 220.

FEATURES

Source

Location/Qualifiers
 1..553
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2376893"
 /clone_lib="Barstead colon HPLRB7"
 /sex="male"
 /dev_stage="adult, age 25"
 /lab_host="DH10B (phage resistant)"
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTACGAATCTGAAGTGGAGCGCGCCGCTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 [5' AATCTACTAGTAAT 3' and 5' ATTACTAGTG 3'], digested
 with Not I and cloned into the Not I and Eco RI sites of
 the modified pT73 vector. Library constructed by Bob
 Barstead."

BASE COUNT

156 a 121 c 97 g 179 t
 ORIGIN

Query Match

Best Local Similarity 1.4%; Score 33; DB 61; Length 553;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2244 CTTTATTAAAAA

ne45b1.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:900285 3',
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

RESULT 30

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

REFERENCE

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

AUTHORS

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

TITLE

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

JOURNAL

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

COMMENT

AA503384/C
 LOCUS
 DEFINITION
 mRNA sequence.
 AA503384 75 bp mRNA EST 19-AUG-1997
 AA503384.1 GI:2238351
 EST.
 human.
 ORGANISM

FEATURES

Source

Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 51.
 Location/Qualifiers
 1..75
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:900285"
 /clone_lib="NCI_CGAP_Co3"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library went through one round of
 normalization."

BASE COUNT 10 a 2 c 1 g 62 t

ORIGIN

Query Match 1.4%; Score 32; DB 34; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.0027;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2245 TTTATTAAAAA
 Db 70 TTTATTAAAAA

RESULT 31

AI960856 160 bp mRNA EST 20-AUG-1999
 SC91e07.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-c1019-733 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AI960856.1 GI:5753569
 soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 eudicotyledons; Magnoliophyta; Magnoliopsida; eudicotyledons; core
 eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 Glycine.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-733"
/clone_lib="Gm-c1019"
/tissue_type="immature seed coats of greenhouse grown
plants"
/lab_host="DH10B (Gibco BRL)"
/notes="vector: pSPORT1 (Life Technologies); Site_1: Not I;
Site_2: Sal I; This cDNA library was constructed from mRNA
isolated from immature seed coats (200-300 mgs) of
greenhouse grown plants. The library was prepared using
the Life Technologies pSuperScript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a Not I restriction
site. Sal I linker adapters were ligated to the
blunt-ended cDNA fragments followed by Not I digestion.
The cDNA fragments were directionally cloned into the Not
I-Sal I restriction site of the pSPORT1 vector. The
ligated cDNA fragments were transformed into E.coli
ElectroMax DH10B host cells (Gibco BRL). This library was
constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      100 a      10 c      16 g      34 t
ORIGIN

Query Match      1.4%; Score 32; DB 63; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTAAAAA.....AAAAAAAAAAAA 2276
|||||.....|
Db 77 TTTATTAAAAA.....AAAAAAAAAAAA 108
|||||.....|

RESULT 32
LOCUS      AU037642      168 bp      mRNA      EST      29-MAR-1999
DEFINITION AU037642 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSE122, mRNA sequence.
ACCESSION  AU037642
VERSION     AU037642
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
ORGANISM    Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1 (bases 1 to 168)
AUTHORS     Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE       The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development.
JOURNAL     DNA Res 5 (6), 335-340 (1998)
MEDLINE     99156227
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2284680.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@akura.cc.tsukuba.ac.jp
PROJECT = "Dictyostelium discoideum cDNA project in Japan".
Location/Qualifiers
1. .168
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSE122"
/clone_lib="dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT      89 a      20 c      6 g      50 t      3 others
ORIGIN

```

```

Query Match      1.4%; Score 32; DB 44; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTAAAAA.....AAAAAAAAAAAA 2276
|||||.....|
Db 131 TTTATTAAAAA.....AAAAAAAAAAAA 162
|||||.....|

RESULT 33
LOCUS      C94407      174 bp      mRNA      EST      15-JUN-1998
DEFINITION C94407 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSK659, mRNA sequence.
ACCESSION  C94407
VERSION     C94407.1 GI:3219022
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
ORGANISM    Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1 (bases 1 to 174)
AUTHORS     Yoshino,R., Morio,T. and Tanaka,Y.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797274.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@akura.cc.tsukuba.ac.jp
PROJECT = "Dictyostelium discoideum cDNA project in Japan".
Location/Qualifiers
1. .174
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSK659"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT      108 a      12 c      8 g      46 t
ORIGIN

Query Match      1.4%; Score 32; DB 41; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTAAAAA.....AAAAAAAAAAAA 2276
|||||.....|
Db 117 TTTATTAAAAA.....AAAAAAAAAAAA 148
|||||.....|

RESULT 34
LOCUS      A1811150/c      178 bp      mRNA      EST      07-JUL-1999
DEFINITION A1811150 NCI_CGAP_Utl1 Homo sapiens cDNA clone IMAGE:2261894 3',
mRNA sequence.
ACCESSION  A1811150
VERSION     A1811150.1 GI:5397716
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 178)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948846.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

```

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

FEATURES
 source
 1. 178
 /location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="2261894"
 /clone_lib="NCI-CGAP_Utl"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH103"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 72 a 24 c 16 g 66 t

ORIGIN

Query Match 1.4%; Score 32; DB 61; Length 178;
 Best Local Similarity 100.0%; Pred. No. 0.0019;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA.....AAAAAAAAAAAAAAAA 2276

Db 32 TTTATTTAAAAA.....AAAAAAAAAAAAAAAA 1

RESULT 35

AW059084

LOCUS

fe24h07.x1 Zebrafish Washu MPIMG EST Danio rerio cDNA 3', mRNA

sequence. 215 bp mRNA EST 29-SEP-1999

DEFINITION

AW059084

VERSION

KEYWORDS

SOURCE

ORGANISM

AW059084.1 GI:5934723

zbrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;

Cyprinoidae; Cyprinidae; Rasbora; Danio.

1 (bases 1 to 215)

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,

Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,

Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,

Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,

Waterston, R., and Wilson, R.

Washu Zebrafish EST Project 1998

Unpublished (1998)

On May 18, 1998 this sequence version replaced gi:3138053.

Contact: Stephen L. Johnson

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrfish@watson.wustl.edu

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

Ressourcen2entrumPrimaDatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7 ET from Amersham.

FEATURES

source

1. 215

/location/Qualifiers

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone_lib="Zebrafish Washu MPIMG EST"

/sex="mixed"

/tissue_type="26 somite embryos, adult livers, shield

stage embryos"

/lab_host="XLI-blue MRF"

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dt)15 primer

[5'pGACTAGTCTAGATCGGAGCGCCCTTTTCTTTT3'];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPORT1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss), adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

BASE COUNT 104 a 24 c 32 g 55 t

ORIGIN

Query Match 1.4%; Score 32; DB 64; Length 215;

Best Local Similarity 100.0%; Pred. No. 0.0018;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA.....AAAAAAAAAAAAAAAA 2276

Db 157 TTTATTTAAAAA.....AAAAAAAAAAAAAAAA 188

RESULT 36

AI685109/c

LOCUS

DEFINITION

AI685109

VERSION

KEYWORDS

SOURCE

ORGANISM

AI685109.1 GI:4896403

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 217)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Jun 5, 1998 this sequence version replaced gi:3187518.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 174.

FEATURES

source

1. 217

/location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2323718"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT      56 a      45 g      81 t
ORIGIN

```

```

Query Match      1.4%; Score 32; DB 50; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2245 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
|||||
Db 42 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 11
|||||

```

RESULT 37

```

AW117010      226 bp mRNA EST 20-OCT-1999
LOCUS      fel6g07.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
DEFINITION      sequence.
ACCESSION      AW117010
VERSION      AW117010.1 GI:6083348
KEYWORDS      EST.
SOURCE      zebrafish.
ORGANISM      Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasbora; Danio.
REFERENCE      1 (bases 1 to 226)
AUTHORS      Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
TITLE      On Jun 5, 1998 this sequence version replaced gi:3189699.
JOURNAL
COMMENT      Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@watson.wustl.edu
CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum Primatardatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: nt7 ET from Amersham.

```

FEATURES

```

source
1..226
Location/Qualifiers
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/notes="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACTAGTCTGATCGAGCGCGCCCTTTTCTTTTCTTTT3'];

```

double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

```

BASE COUNT      113 a      26 c      31 g      56 t
ORIGIN

```

```

Query Match      1.4%; Score 32; DB 69; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2245 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276
|||||
Db 157 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 188
|||||

```

RESULT 38

```

AU029417      237 bp mRNA EST 19-OCT-1998
LOCUS      AU029417 Rice panicle shorter than 3cm Oryza sativa cDNA clone
DEFINITION      E30335_62, mRNA sequence.
ACCESSION      AU029417
VERSION      AU029417.1 GI:3762665
KEYWORDS      EST.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE      1 (bases 1 to 237)
AUTHORS      Sasaki, T. and Yamamoto, K.
TITLE      Rice cDNA from panicle
JOURNAL      Unpublished (1997)
COMMENT      On Jan 17, 1998 this sequence version replaced gi:1899911.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT "RGP".

```

FEATURES

```

source
1..237
Location/Qualifiers
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="E30335_62"
/dev_stage="shorter than 3cm"
/notes="Organ: panicle"
BASE COUNT      87 a      41 c      69 t      2 others
ORIGIN

```

```

Query Match      1.4%; Score 32; DB 43; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


2245 TTTATTTAAAAA..... 2276
 |||||||.....
 197 TTTATTTAAAAA..... 228

SULT 39
 571665
 CCUS
 FFINATION
 CSSION
 SSION
 YWORDS
 URCE
 ORGANISM

AA571665 243 bp mRNA EST 27-AUG-1997
 vm09e04_r1 Knowles Solter mouse blastocyst B1 Mus musculus CDNA
 clone IMAGE:989694 5', mRNA sequence.
 AA571665
 AA571665.1 GI:2346594
 EST.
 house mouse.
 Mus musculus
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 243)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisler,S., Kucab,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On May 9, 1995 this sequence version replaced gi:802596.
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL : contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:361974
 Putative full length read
 vector to vector length is 632
 High quality sequence stop: 210.
 Location/Qualifiers
 1. .243
 /organism="Mus musculus"
 /strain="B6D2 Fl/J"
 /db_xref="taxon:10090"
 /clone_lib="Knowles Solter mouse blastocyst B1"
 /clone_image="IMAGE:989694"
 /issue_type="blastocyst"
 /dev_stage="embryo (pre-implantation)"
 /lab_host="DH10B"
 /note="Organ: embryo; Vector: pSPORT; Site: 1: NotI;
 Site 2: SalI: Cloned unidirectionally from mRNA prepared
 from 800 blastocysts Primer: SalI(df):
 5'-CGGTGACGTCGACCGCTTTTITTTT-3'. cDNAs were
 cloned into the NotI/SalI sites of a pSPORT vector (Life
 Technologies). Two different size selections: B1 (large
 inserts) and B3."
 142 a 30 c 43 g 28 t
 IGIN

Query Match 1.4%; Score 32; DB 35; Length 243;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0

2245 TTTATTTAAAAA..... 2276
 |||||||.....
 120 TTTATTTAAAAA..... 151

SULT 40
 038878
 CCUS
 FFINATION
 CSSION
 SSION
 YWORDS
 URCE
 ORGANISM

AA038878 243 bp mRNA EST 29-MAR-1999
 AU038878 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium

```

Seq primer: -40UP from Gibco
High quality sequence stop: 245.
Location/Qualifiers
1. .247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Brn35"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.33 kb. Tumor types include:
meningioma, oligodendroglioma, astrocytoma (grade II),
medulloblastoma, astrocytoma (grade IV). Life Technologies
catalog #: 11544-012"
BASE COUNT      85 a  28 c  27 g  107 t
ORIGIN

Query Match      1.4%; Score 32; DB 64; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA 2276
|||||
Db 33 TTTATTTAAAAA 2276
|||||

RESULT 42
LOCUS      C90433      256 bp      mRNA      EST      20-APR-1998
DEFINITION C90433 Dictyostelium discoideum SS (H.Urushi-hara) Dictyostelium
ACCESSION  C90433
VERSION    C90433.1 GI:3060053
KEYWORDS   EST
SOURCE     Dictyostelium discoideum.
ORGANISM   Eukaryota; Dictyostelid; Dictyostelium.
REFERENCE  Yoshino,R., Morio,T. and Tanaka,Y.
AUTHORS    Developmental cDNA in Dictyostelium discoideum
TITLE       Unpublished (1997)
JOURNAL    On Jan 19, 1998 this sequence version replaced gi:2151596.
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402huesakura.cc.tsukuba.ac.jp.
            Location/Qualifiers
1. .256
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SSI480"
/dev_stage="slug"
BASE COUNT      150 a  25 c  18 g  63 t
ORIGIN

Query Match      1.4%; Score 32; DB 40; Length 256;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA 2276
|||||
Db 202 TTTATTTAAAAA 233
|||||

RESULT 43

```

```

AW195999/c
LOCUS      AW195999      265 bp      mRNA      EST      29-NOV-1999
DEFINITION x186c05.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2681576 3',
            mRNA sequence.
ACCESSION  AW195999
VERSION    AW195999.1 GI:6475229
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 265)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137479.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 247.
Location/Qualifiers
1. .265
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT      74 a  37 c  51 g  103 t
ORIGIN

Query Match      1.4%; Score 32; DB 74; Length 265;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA 2276
|||||
Db 48 TTTATTTAAAAA 17
|||||

RESULT 44
LOCUS      AW1872568/c      269 bp      mRNA      EST      30-AUG-1999
DEFINITION wm76g11.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2441924 3',
            similar to contains element MSRI repetitive element ;, mRNA
            sequence.
ACCESSION  AW1872568
VERSION    AW1872568.1 GI:5546617
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 269)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT     On Dec 20, 1995 this sequence version replaced gi:1134419.

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 266.
Location/Qualifiers
1. .269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2441924"
/clone_lib="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 70 a 36 c 53 g 110 t
ORIGIN
Query Match 1.4%; Score 32; DB 62; Length 269;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2245 TTTATTTAAAAA... 2276
Db 43 TTTATTTAAAAA... 12

RESULT 45
AA749402/c
LOCUS AA749402 272 bp mRNA EST 18-FEB-1998
DEFINITION ny04h08.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270815 3', mRNA sequence.
ACCESSION AA749402
VERSION AA749402.1 GI:2789360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 272)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gi:1877690.
Contact: Robert Strausberg, Ph.D.

REFERENCE
AUTHORS Tel: (301) 496-1550
TITLE Email: Robert.Strausberg@nih.gov
JOURNAL Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
COMMENT cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1288 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham

FEATURES
Source
High quality sequence stop: 170.
Location/Qualifiers
1. .272
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1270815"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dt) primer
[5'-TGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 25 c 30 g 119 t
ORIGIN

Query Match 1.4%; Score 32; DB 38; Length 272;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2245 TTTATTTAAAAA... 2276
Db 39 TTTATTTAAAAA... 8

Search completed: April 6, 2000, 13:17:57
Job time: 21662 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:18:09 ; Search time 6421.81 Seconds
(without alignments)
79.372 Million cell updates/sec

Title: US-09-090-672B-9
Perfect score: 135
Sequence: 1 TTCTGACATGACTAAGAG.....AGGGACTGGACTAAATAGC 135

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :		EST:*	
		1: em_est1:*	
		2: em_est2:*	
		3: em_est3:*	
		4: em_est4:*	
		5: em_est5:*	
		6: em_est6:*	
		7: em_est7:*	
		8: em_est8:*	
		9: em_est9:*	
		10: em_est10:*	
		11: em_est11:*	
		12: em_est12:*	
		13: em_est13:*	
		14: em_est14:*	
		15: em_est15:*	
		16: em_est16:*	
		17: em_est17:*	
		18: em_est18:*	
		19: em_est19:*	
		20: gb_est1:*	
		21: gb_est2:*	
		22: gb_est3:*	
		23: gb_est4:*	
		24: gb_est5:*	
		25: gb_est6:*	
		26: gb_est7:*	
		27: gb_est8:*	
		28: gb_est9:*	
		29: gb_est10:*	
		30: gb_est11:*	
		31: gb_est12:*	
		32: gb_est13:*	
		33: gb_est14:*	
		34: gb_est15:*	
		35: gb_est16:*	
		36: gb_est17:*	
		37: gb_est18:*	
		38: gb_est19:*	
		39: gb_est20:*	
		40: gb_est21:*	
		41: gb_est22:*	
		42: gb_est23:*	
		43: gb_est24:*	

44: gb_est25:*	
45: gb_est26:*	
46: gb_est27:*	
47: gb_est28:*	
48: gb_est29:*	
49: gb_est30:*	
50: gb_est31:*	
51: gb_est32:*	
52: em_est20:*	
53: em_est21:*	
54: em_est22:*	
55: em_est23:*	
56: em_est24:*	
57: em_est25:*	
58: em_est26:*	
59: gb_est33:*	
60: gb_est34:*	
61: gb_est35:*	
62: gb_est36:*	
63: gb_est37:*	
64: gb_est38:*	
65: em_est27:*	
66: em_est28:*	
67: em_est29:*	
68: em_est30:*	
69: gb_est39:*	
70: gb_est40:*	
71: gb_est41:*	
72: gb_est42:*	
73: gb_est43:*	
74: gb_est44:*	
75: em_est31:*	
76: em_est32:*	
77: em_est33:*	
78: em_est34:*	
79: gb_gss1:*	
80: gb_gss2:*	
81: gb_gss3:*	
82: gb_gss4:*	
83: em_gss1:*	
84: em_gss2:*	
85: em_gss3:*	
86: em_gss4:*	
87: gb_gss5:*	
88: gb_gss6:*	
89: gb_gss7:*	
90: gb_gss8:*	
91: gb_gss9:*	
92: em_gss5:*	
93: em_gss6:*	
94: em_gss7:*	
95: em_gss8:*	
96: em_gss9:*	
97: em_gss10:*	
98: em_gss11:*	
99: gb_gss10:*	
100: gb_gss11:*	
101: em_gss12:*	
102: gb_gss12:*	
103: gb_gss13:*	
104: gb_gss14:*	
105: gb_gss15:*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		Description	
No.	Score	Match	Length	ID	
c 1	18	13.3	278	41	AI007840
					AI007840 EST202291

```
C 2 18 13.3 364 42 A1145820
C 3 18 13.3 447 48 A1548491
C 4 18 13.3 498 43 A1100865
C 5 18 13.3 534 43 A1101330
C 6 18 13.3 903 100 A0271832
C 7 17 12.6 175 81 B26945
C 8 17 12.6 340 30 A0072249
C 9 17 12.6 356 30 A0261914
C 10 17 12.6 404 24 H73595
C 11 17 12.6 414 38 A0767354
C 12 17 12.6 427 28 A0126408
C 13 17 12.6 429 99 A0184537
C 14 17 12.6 435 30 A0262551
C 15 17 12.6 447 30 A0228289
C 16 17 12.6 451 40 A0935611
C 17 17 12.6 468 74 A0221275
C 18 17 12.6 476 35 A0586533
C 19 17 12.6 483 99 A0240973
C 20 17 12.6 484 102 A0402549
C 21 17 12.6 523 48 A1612731
C 22 17 12.6 528 69 A0106289
C 23 17 12.6 533 100 A0277193
C 24 17 12.6 577 105 A0657152
C 25 17 12.6 601 81 B29489
C 26 17 12.6 618 100 A0278681
C 27 17 12.6 620 91 A0157444
C 28 17 12.6 764 90 A0078191
C 29 17 12.6 766 82 A0738602
C 30 17 12.6 931 81 B09381
C 31 16 11.9 96 27 A0052408
C 32 16 11.9 173 44 A1278373
C 33 16 11.9 195 49 A0111313
C 34 16 11.9 213 35 AA570404
C 35 16 11.9 262 44 A1277304
C 36 16 11.9 263 105 A0663218
C 37 16 11.9 272 20 T02730
C 38 16 11.9 300 35 C34963
C 39 16 11.9 300 35 C36512
C 40 16 11.9 300 35 C51747
C 41 16 11.9 300 35 C52581
C 42 16 11.9 300 35 C52912
C 43 16 11.9 300 62 AV178626
C 44 16 11.9 300 62 AV184617
C 45 16 11.9 305 72 AV332351
```

ALIGNMENTS

```
RESULT 1
LOCUS A1007840/c
DEFINITION EST202291 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
ACCESSION BBRAV56.3 end, mRNA sequence.
VERSION A1007840.1 GI:3221672
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE Lee.N.H., Glodek.A., Chandra.I., Mason.T.M., Quackenbush.J.,
Kerlavage.A.R. and Adams.M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
JOURNAL On Jan 17, 1998 this sequence version replaced gi:1900158.
COMMENT Contact: Lee, NH
ATCC The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
```

```
FEATURES
source
location/Qualifiers
1..278
/organism="Rattus sp."
/db_xref="ATCC (inhost):2016112"
/db_xref="taxon:10118"
/clone="BBRAV56"
/note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 86 a 70 c 53 g 59 t
ORIGIN
Query Match 13.3%; Score 18; DB 41; Length 278;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 CAAGGGACTGGACTAAAT 131
Db 241 CAAGGGACTGGACTAAAT 224
|||||
RESULT 2
LOCUS A1145820/c
DEFINITION A1145820 364 bp mRNA EST 05-JUL-1999
UI-R-BT0-ql-g-12-0-UI.sl UI-R-BT0 Rattus norvegicus cDNA clone
ACCESSION A1145820
VERSION A1145820.1 GI:3667619
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE Bonaldo.M.F., Lennon.G. and Soares.M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT On Jan 17, 1998 this sequence version replaced gi:2152861.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult NR6thalamus library. CDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (infoimage.llnl.gov). IMAGE
ID-1788870 The following repetitive elements were found in this
cDNA sequence: 1-27, >AT-rich#Low_complexity
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1..364
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BT0-ql-g-12-0-UI"
/clone_lib="UI-R-BT0"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT73B-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
(UI-R-BT0) consists of a mixture of individually tagged
```

normalized libraries constructed from rat hippocampus, thalamus, mid-brain, medulla, corpus striatum, cerebral cortex and testis. The tag used to identify the source tissue is a string of 3-6 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. This library was then subcloned using a driver consisting of a mixture of all clones from UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and UI-R-C2p."

BASE COUNT 106 a 89 c 66 g 103 t
ORIGIN

Query Match 13.3%; Score 18; DB 42; Length 364;

Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGACTGGACTAAAT 131
|||||
DB 258 CAAGGACTGGACTAAAT 241

RESULT 3

AI548491/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

On Jun 5, 1998 this sequence version replaced gi:3188189.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dr track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dr track served to identify it as a clone from the

normalized brain library cDNA Library Preparation: M.B. Soares Lab

Clone distribution: clones will be available through Research

Genetics (www.resgen.com) The following repetitive elements were

found in this cDNA sequence: 1-34, >AT-rich#low_complexity

Seq primer: M13 Forward.

Location/Qualifiers

1. .447

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-C3-to-a-04-0-UI"

/clone_lib="UI-R-C3"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C3

library is a subcloned library of a series, ultimately

derived from a mixture of individually tagged normalized

libraries from rat placenta, adult lung, brain, liver,

kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day embryos, after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subcloned libraries were generated in this process: UI-R-C3, UI-R-C2p, UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subcloned library (UI-R-C3) was constructed as follows: PCR amplified cDNA inserts from UI-R-C2p clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C2p library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C3 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)".

BASE COUNT 134 a 110 c 81 g 122 t
ORIGIN

Query Match 13.3%; Score 18; DB 48; Length 447;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGACTGGACTAAAT 131
|||||
DB 265 CAAGGACTGGACTAAAT 248

RESULT 4

AI100865/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)

Other ESTs: TC54489

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@igr.org

Seq primer: M13-21.

Location/Qualifiers

1. .498

/organism="Rattus sp."

/db_xref="ATCC (inhost):2023824"

/db_xref="taxon:10118"

/clone="RBBB71"

/clone_lib="Normalized rat brain, Bento Soares"

/note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;

Site_2: NotI"

BASE COUNT 149 a 120 c 108 g 121 t

ORIGIN

```

Query Match      13.3%; Score 18; DB 43; Length 498;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGGACTGGACTAAAT 131
    |||||
Db 247 CAAGGGACTGGACTAAAT 230

RESULT 5
AII01330/c
LOCUS      AII01330      534 bp      mRNA      EST      31-JAN-1999
DEFINITION EST210519 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRBL55 3' end, mRNA sequence.
ACCESSION  AII01330
VERSION     AII01330.1 GI:3706248
KEYWORDS    Rattus sp.
SOURCE      Rattus sp.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 534)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
JOURNAL     Unpublished (1998)
COMMENT     Other_ESTS: TC54489
            Contact: Lee, NH
            ATCC
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@stg.org
            Seq primer: M13-21.

FEATURES             Location/Qualifiers
     source            1..534
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2029307"
                     /db_xref="taxon:10118"
                     /clone="RBRBL55"
                     /clone_lib="Normalized rat brain, Bento Soares"
                     /note="Organ: brain; Vector: pT73Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT  155 a 130 c 117 g 132 t

Query Match      13.3%; Score 18; DB 43; Length 534;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CRAAGGACTGGACTAAAT 131
    |||||
Db 244 CRAAGGACTGGACTAAAT 227

RESULT 6
AQ271832/c
LOCUS      AQ271832      903 bp      DNA      GSS      03-NOV-1998
DEFINITION nbxb0026D08f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0026D08f, genomic survey sequence.
ACCESSION  AQ271832
VERSION     AQ271832.1 GI:3825147
KEYWORDS    GSS.
SOURCE      Oryza sativa.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE  1 (bases 1 to 903)
AUTHORS    Wing,R.A. and Dean,R.A.

```

```

TITLE      A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL     Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 88.
            Location/Qualifiers
     source            1..903
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0026D08f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world population especially those
                     inhabiting highly populated areas of the humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991). The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to facilitate positional cloning, physical mapping
                     and genome sequencing of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents.
                     The deep coverage allows the isolation a particular
                     sequence with a probability of 99.9 %. Two high density
                     filters, each containing 18,432 clones (doubly spotted),
                     represent the whole library for colony screening."
BASE COUNT  192 a 183 c 89 g 419 t 20 others
ORIGIN

Query Match      13.3%; Score 18; DB 100; Length 903;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AATGAGTAAGAGAGAGA 25
    |||||
Db 750 AATGAGTAAGAGAGAGA 733

RESULT 7
B26945
LOCUS      B26945      175 bp      DNA      GSS      13-OCT-1997
DEFINITION T24A9TR TAMU Arabidopsis thaliana genomic clone T24A9, genomic
            survey sequence.
ACCESSION  B26945
VERSION     B26945.1 GI:2512911
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
            Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsids.
REFERENCE  1 (bases 1 to 175)
AUTHORS    Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
            Venter,J.C.
            Use of a BAC End Sequence Database To Identify Minimal Overlaps for
            Arabidopsis Genomic Sequencing
TITLE

```


JOURNAL
COMMENT

Unpublished (1997)
Other GSSs: T2A9TfB
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 175.
Location/Qualifiers
1. .175

FEATURES

source

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T2A9"
/clone_lib="TAMU"
/sex="hermaphrodite"
/note="vector: BelOBACII; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"
60 a 31 c 25 g 59 t

BASE COUNT
ORIGIN

Query Match 12.6%; Score 17; DB 81; Length 175;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ATTAAGATTTCATFAG 60
|||||

DB 116 ATTAAGATTTCATFAG 132
|||||

RESULT 8

AQ072249

LOCUS

DEFINITION
sapiens genomic clone Plate=3024 Col=1 Row=N, genomic survey
sequence.

ACCESSION

AQ072249

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

1 (bases 1 to 340)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL

MEDLINE

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3024 row: N column: 1

Class: BAC ends

High quality sequence stop: 340.

Location/Qualifiers

1. .340

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate=3024 Col=1 Row=N"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

FEATURES

source

Query Match 12.6%; Score 17; DB 30; Length 356;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCITGCCCTTTGGTAT 45
|||||

DB 354 TCITGCCCTTTGGTAT 338
|||||

/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 127 a 44 c 58 g 111 t
ORIGIN

Query Match 12.6%; Score 17; DB 90; Length 340;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ATTATCATGAGCAAT 66
|||||

DB 64 ATTATCATGAGCAAT 80
|||||

RESULT 9

AA261914/c

LOCUS

DEFINITION
AA261914 356 bp mRNA EST 13-AUG-1997
ZS17dl1.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:685461 3',
mRNA sequence.

ACCESSION

AA261914

VERSION

AA261914.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 8, 1995 this sequence version replaced gi:800902.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -41m13 fwd. ET from Amersham.

Location/Qualifiers

1. 356

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:685461"

/clone_lib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

15'-TGTACCAATCTCAAGTGGAGCGCGCTCATTTTTTTTTTTT-

3'. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 117 c 88 g 66 t

ORIGIN

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 427)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chispeilli, B.,
Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marita, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 9704478
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397564.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 from Amersham
High quality sequence stop: 277.
Location/Qualifiers
1. .427
/organism="Homo sapiens"
/db_xref="GDB:3844042"
/db_xref="taxon:9606"
/clone="IMAGE:511547"
/lab_host="Stratagene colon (#937204)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'

BASE COUNT 114 a 121 c 114 g 76 t 2 others
ORIGIN

Query Match 12.6%; Score 17; DB 28; Length 427;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 25 AGGCTCTTGCCTTGG 41
|||||
Db 354 AGGCTCTTGCCTTGG 338

RESULT 13
AQ184537/c
LOCUS
DEFINITION HS.2205.AL.G01.MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2205 Col-1 Row-M, genomic survey
sequence.
ACCESSION AQ184537
VERSION AQ184537.1 GI:3583979
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 429)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2205 row: M column: 1
Class: BAC ends
High quality sequence stop: 429.
Location/Qualifiers
1. .429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2205 Col=1 Row=M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 126 a 82 c 79 g 139 t 3 others
ORIGIN

Query Match 12.6%; Score 17; DB 99; Length 429;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 AGAAGAAGAGGGTCTT 32
|||||
Db 155 AGAAGAAGAGGGTCTT 139

RESULT 14
AA262551
LOCUS
DEFINITION zs17d11.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:685461 5',
mrna sequence.
ACCESSION AA262551
VERSION AA262551.1 GI:1897913
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 435)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 8, 1995 this sequence version replaced gi:801104.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read
The vector to vector length is 436
Seq primer: -28m13 rev2 ET from Amersham.
Location/Qualifiers
1. .435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:685461"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGGCTCATTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 112 c 136 g 104 t
ORIGIN

Query Match 12.6%; Score 17; DB 30; Length 435;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCTTGCCCTTGGTTAT 45
|||||

Db 72 TCTTGCCCTTGGTTAT 88

RESULT 15

AA228289
LOCUS nc38c02.r1 NCI_CGAP_Pr2 Homo sapiens CDNA clone IMAGE:1010402
DEFINITION similar to contains Alu repetitive element; contains element PTR5 repetitive element ; mRNA sequence.

ACCESSION AA228289
VERSION AA228289.1 GI:1849850
KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 447)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1393543.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuauqui, M.D., Michael Emerit-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

Cloning strategy: Washington University Genome Sequencing Center
Cloning distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 505 Std Error: 0.00

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 409.

FEATURES

source

1..447

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1010402"

/clone_lib="NCI_CGAP_Pr2"

/sex="Male"

/dev_stage="45 years old"

/lab_host="DH10B"

/note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from

5,000-10,000 microdissected preneoplastic cells histologically-determined to be prostatic intraepithelial neoplasia 2 (PIN2) cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting

PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 128 a 96 c 111 g 112 t
ORIGIN

Query Match 12.6%; Score 17; DB 30; Length 447;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GACATGAGTAAGAAGA 21
|||||

Db 61 GACATGAGTAAGAAGA 77

RESULT 16

AA935611
LOCUS AA935611 451 bp mRNA EST 07-JUL-1998
DEFINITION Op26g10.s1 Soares_NFL_T_GBC_S1 Homo sapiens CDNA clone IMAGE:1578018 3' similar to WP:C14C11.3 CE06822 ; mRNA sequence.

ACCESSION AA935611
VERSION AA935611.1 GI:3092768
KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 451)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:1900158.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1658 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 148.

FEATURES

source

1..451

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1578018"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI_CGAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239, 728408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 106 a 121 c 126 g 98 t
ORIGIN

Query Match 12.6%; Score 17; DB 40; Length 451;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAATGAGTAAGAAGA 23
|||||

Db 241 CAATGAGTAAGAAGA 257

```

RESULT 17
AW221275      468 bp      mRNA      EST      07-DEC-1999
LOCUS      EST297744 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION      cDNA cLEF13N5, mRNA sequence.
ACCESSION      AW221275
VERSION      AW221275.1 GI:65332959
KEYWORDS      EST.
SOURCE      tomato.
ORGANISM      Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Potatoe; Lycopersicon.
1 (bases 1 to 468)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
On Mar 10, 1998 this sequence version replaced gi:2948814.
Contact: David Frisch
Clemson University
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES
Source      Location/Qualifiers
1..468
/cultivar="TA96"
/organism="Lycopersicon esculentum"
/db_xref="taxon:4081"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLEF - Fruit were tagged at the lcm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT      136 a 108 c 92 g 132 t
ORIGIN
Query Match      12.6%; Score 17; DB 74; Length 468;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 ACAGCAGCAGAGATTAG 103
Db 178 ACCAGCAGCAGAGATTAG 194
|||||
RESULT 18
AA586533      476 bp      mRNA      EST      12-SEP-1997
LOCUS      nd46c09.s1 NCI-CGAP_Col Homo sapiens cDNA clone IMAGE:816016.3/
DEFINITION      similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE,
LIVER (HUMAN);, mRNA sequence.
ACCESSION      AA586533
VERSION      AA586533.1 GI:2397347
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 9, 1995 this sequence version replaced gi:802594.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -21ml3 forward (Amersham)
High quality sequence stop: 1.
FEATURES
Source      Location/Qualifiers
1..476
/organism="Homo sapiens"
/db_xref="GDB:6037392"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Col"
/tissue_type="bulk tumor"
/lab_host="DH10B"
/Note="Organ: colon; Vector: pCMV-SPORT2; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt."
BASE COUNT      106 a 114 c 134 g 122 t
ORIGIN
Query Match      12.6%; Score 17; DB 35; Length 476;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 AAATCGTGTATACCA 90
Db 253 AAATCGTGTATACCA 269
|||||
RESULT 19
AQ240973      483 bp      DNA      GSS      21-APR-1999
LOCUS      RPC111-64B24.TK RPC1-11 Homo sapiens genomic clone RPC1-11-64B24,
DEFINITION      genomic survey sequence.
ACCESSION      AQ240973
VERSION      AQ240973.1 GI:3668557
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 483)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPC111-64B24.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamad@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong

```

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..483

/organism="Homo sapiens"

/db_xref="GDB:7524239"

/db_xref="taxon:9606"

/clone="RPCI-11-64B24"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

153 a 96 c 92 g 142 t

BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; DB 99; Length 483;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 TTATCATAGCAATAA 68

Db 172 TTATCATAGCAATAA 156

RESULT 20

AQ402549/c

LOCUS

DEFINITION AQ402549 484 bp DNA GSS 13-MAR-1999

HS_5054_A1_E06_47A RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plasmid-630 Col-11 Row-I, genomic survey sequence.

ACCESSION AQ402549

VERSION AQ402549.1

KEYWORDS GSS

SOURCE human.

ORGANISM

REFERENCE

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)or from Research Genetics (<http://inforesgen.com>). BAC end Web Server:<http://www.hsc.washington.edu>

Plate: 630 Row: I Column: 11

Seq primer: T7

Class: BAC ends

High quality sequence stop: 484.

Location/Qualifiers

1..484

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-630 Col-11 Row-I"

/clone_lib="RPCI-11 Human Male BAC Library"

FEATURES

source

Location/Qualifiers

1..484

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-630 Col-11 Row-I"

/clone_lib="RPCI-11 Human Male BAC Library"

/sex="male"
/note="Vector: pBACE3.6; Genomic sequence of BAC ends"
144 a 101 c 105 g 128 t 6 others

BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; DB 102; Length 484;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ACAATAAACCAAGGGA 120

Db 74 ACAATAAACCAAGGGA 58

RESULT 21

AI612731/c

LOCUS

DEFINITION

AI612731

523 bp mRNA

EST 21-APR-1999

tp11b09.x1 NCI-CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2187449

similar to gb:X06617.40S RIBOSOMAL PROTEIN S11 (HUMAN);, mRNA

sequence.

ACCESSION AI612731

VERSION AI612731.1

KEYWORDS EST

SOURCE human.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 18, 1998 this sequence version replaced gi:3136494.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Suck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 67.

Location/Qualifiers

1..523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2187449"

/clone_lib="NCI-CGAP_Gas4"

/tissue_type="poorly differentiated adenocarcinoma with

signal ring cell features"

/lab_host="DH10B"

/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 1.69 kb. Life Technologies catalog #:

11549-011"

104 a 126 c 138 g 155 t

BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; DB 48; Length 523;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AAGAAAGAGGCTTGC 34

Db 507 AAGAAAGAGGCTTGC 491

```

RESULT 22
AW106269
LOCUS
DEFINITION
IMAGE:2225591 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
TITLE
The WashU-NCI Mouse EST Project 1999
JOURNAL
COMMENT
On Dec 20, 1995 this sequence version replaced gi:1134183.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousest@watson.wustl.edu
This clone is available royalty-free through LLMU; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1005787
Seq primer: custom primer used
High quality sequence stop: 443.
FEATURES
Location/Qualifiers
1..528
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2225591"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTC);
Site_2: DraIII (CACTGTGTC); 1st strand cDNA was primed
with an oligo(dT) primer [ATGCGCTTTTITTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGGCGCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTC, 3' site
CACTGTGTC). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCGCTGCTAAAGCTGCG and 3' end primer
CGACCTGACGCTGAGCACA."
BASE COUNT 155 a 107 c 112 g 154 t
ORIGIN

Query Match 12.6%; Score 17; DB 69; Length 528;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GTAAGAAGAAGAGGCT 29
Db 369 GTAAGAAGAAGAGGCT 385

RESULT 23
AQ277193
LOCUS
DEFINITION
IMAGE:2522C17.TF C17B1-E1 Homo sapiens genomic clone 2522C17,
genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.
TITLE
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building (1998)
JOURNAL
COMMENT
Unpublished (1998)
Other_GSSs: CITBI-E1-2522C17.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamad@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Location/Qualifiers
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2522C17"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltch Human BAC Library D"
BASE COUNT 143 a 136 c 176 g 176 t
ORIGIN

Query Match 12.6%; Score 17; DB 100; Length 533;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAATGAGTAGAAGAAA 23
Db 397 CAATGAGTAGAAGAAA 413

RESULT 24
AQ657152/c
LOCUS
DEFINITION
IMAGE:12F21.TF Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-12F21, genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
AUTHORS
El-Sayed, N., Zhao, S., Zhao, H., Gull, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,
Donelson, J., Fraser, C. and Adams, M.
TITLE
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
JOURNAL
COMMENT
Unpublished (1999)
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

```

Tel: 301 838 0200
Fax: 301 838 0208

Email: nelsaved@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.

FEATURES

Location/Qualifiers

1..577
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-12F21"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 192 a 97 c 168 g 120 t
ORIGIN

Query Match 12.6%; Score 17; DB 105; Length 577;

Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 AGATTATCATGACGA 64

Db 87 AGATTATCATGACGA 71

RESULT 25

B29489
LOCUS
DEFINITION B29489 601 bp DNA GSS 13-OCT-1997
F26010TFB IGF Arabidopsis thaliana genomic clone F26010, genomic survey sequence.
ACCESSION B29489
VERSION B29489.1 GI:2515455
KEYWORDS GSS.
SOURCE thale cress.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eutosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 601)

Rounsley, S.D., Kelley, J.M., Field, C.E., Craven, M.B., Adams, M.D. and Venter, J.C.

Use of a BAC End Sequence Database To Identify Minimal Overlaps for Arabidopsis Genomic Sequencing

JOURNAL

COMMENT
Other_GSSs: F26010TRC
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 601.

FEATURES

Location/Qualifiers

1..601

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F26010"
/clone_lib="IGF"
/sex="hermaphrodite"

/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI; Produced by Thomas Altmann"
BASE COUNT 213 a 94 c 105 g 189 t
ORIGIN

Query Match 12.6%; Score 17; DB 81; Length 601;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ATTAAGATTATCATAG 60

Db 581 ATTAAGATTATCATAG 597

RESULT 26

AQ278681
LOCUS
DEFINITION AQ278681 618 bp DNA GSS 22-NOV-1998
CITBI-EI-251685.TR CITBI-EI Homo sapiens genomic clone 251685, genomic survey sequence.

ACCESSION AQ278681

VERSION AQ278681.1 GI:3904649

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 618)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready Map Building

Unpublished (1998)

Other_GSSs: CITBI-EI-251685.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES

Location/Qualifiers

1..618

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="251685"

/clone_lib="CITBI-EI"

/sex="male"

/cell_type="sperm"

/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;

Caltech Human BAC Library D"

BASE COUNT 173 a 153 c 127 g 165 t

ORIGIN

Query Match 12.6%; Score 17; DB 100; Length 618;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AGTAAGAAGAAAGAGG 28

|||||


```

Db 330 ACTAAGAGAAAGAGGG 346

RESULT 27
AQ157444
LOCUS
DEFINITION
  AQ157444 620 bp DNA GSS 12-SEP-1998
  nbxb0009B13r CUGI Rice BAC Library Oryza sativa genomic clone
  nbxb0009B13r, genomic survey sequence.
ACCESSION
  AQ157444
VERSION
  AQ157444.1 GI:3554469
KEYWORDS
  GSS.
SOURCE
  Oryza sativa.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
  Poaceae; Oryza.
  1 (bases 1 to 620)
REFERENCE
  Wing, R.A. and Dean, R.A.
  A BAC End Sequencing Framework to Sequence the Rice Genome
  Unpublished (1998)
  Contact: Wing RA
  Clemson University Genomics Institute
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: GGAACAGCTATGACCATG
  Class: BAC ends
  High quality sequence start: 54
  High quality sequence stop: 294.
FEATURES
  source
  1..620
  Location/Qualifiers
    /organism="Oryza sativa"
    /strain="Japonica"
    /cultivar="Nipponbare"
    /db_xref="taxon:4530"
    /clone_lib="CUGI Rice BAC Library"
    /tissue_type="Leaf"
    /lab_host="E. coli DH10B"
    /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
    HindIII; Rice is one of two most popular grains in the
    world. Half of the world population especially those
    inhabiting highly populated areas of the humid tropics
    and subtropics, rely on rice as their primary source of
    carbohydrate. Monocotyledonous rice is a diploid plant
    (2n=24) with a haploid genome equivalent of 431 Mbp
    (Arumuganathan and Earle, 1991). The relatively small
    genome of rice, three times larger than that of
    Arabidopsis, makes it suitable for genomic studies. In
    order to facilitate positional cloning, physical mapping
    and genome sequencing of rice, we have constructed a BAC
    library from Oryza sativa, Nipponbare variety. The
    library contains 36,864 clones with an average insert size
    of 128.5 kb providing 10.9 haploid genome equivalents.
    The deep coverage allows the isolation a particular
    sequence with a probability of 99.9 %. Two high density
    filters, each containing 18,432 clones (doubly spotted),
    represent the whole library for colony screening."
  BASE COUNT 174 a 99 c 104 g 243 t
  ORIGIN

Query Match 12.6%; Score 17; DB 91; Length 620;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 TCCTGCCCTTGGTTAT 45
|||||
Db 560 TCCTGCCCTTGGTTAT 576

```

```

RESULT 28
AQ078191/c
LOCUS
DEFINITION
  AQ078191 764 bp DNA GSS 20-AUG-1998
  CIT-HSP-2360M3.TR CIT-HSP Homo sapiens genomic clone 2360M3,
  genomic survey sequence.
ACCESSION
  AQ078191
VERSION
  AQ078191.1 GI:3439375
KEYWORDS
  GSS.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 764)
REFERENCE
  Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
  Venter,J.C.
  Use of a random human BAC End Sequence Database for Sequence-Ready
  Map Building
  Unpublished (1998)
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: M13 Reverse
  Class: BAC ends.
FEATURES
  source
  1..764
  Location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="CIT-HSP"
    /sex="Male"
    /cell_type="Sperm"
    /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
    HindIII"
  BASE COUNT 241 a 174 c 143 g 206 t
  ORIGIN

Query Match 12.6%; Score 17; DB 90; Length 764;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 TTATCATAGAGCAATAA 68
|||||
Db 82 TTATCATAGAGCAATAA 66

RESULT 29
AQ738602/c
LOCUS
DEFINITION
  AQ738602 766 bp DNA GSS 16-JUL-1999
  HS_5381.B1.B07.SP6 RPCI-11 Human Male BAC Library Homo sapiens
  genomic clone Plate-957 Col-13 Row-D, genomic survey sequence.
ACCESSION
  AQ738602
VERSION
  AQ738602.1 GI:5516124
KEYWORDS
  GSS.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 766)
REFERENCE
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  JOURNAL

```

```

MEDLINE
COMMENT
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact pieter de Jong
(pieter@edj.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.husc.washington.edu
Plate: 957 row: D column: 13
Seq primer: Sp6
Class: BAC ends
High quality sequence stop: 766.
FEATURES
Source
1..766
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 160 a 251 c 179 g 176 t
ORIGIN
Query Match 12.6%; Score 17; DB 82; Length 766;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GAGTAGAAGAAGAGG 27
|||||
Db 724 GAGTAGAAGAAGAGG 708

RESULT 30
B09381
LOCUS
DEFINITION
T24A9-Sp6 TAMU Arabidopsis thaliana genomic clone T24A9, genomic
survey sequence.
ACCESSION B09381
VERSION B09381.1 GI:2090501
KEYWORDS
SOURCE
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 931)
Feng J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J. Sequences at ATGC
BAC End Sequences at ATGC
Other_GSSs: T24A9-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 85
High quality sequence stop: 307.
FEATURES
Source
1..931
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="T24A9"
/sex="hermaphrodite"
/note="Vector: BelOBAcII; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"
BASE COUNT 283 a 190 c 154 g 298 t 6 others
ORIGIN
Query Match 12.6%; Score 17; DB 81; Length 931;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ATTAAGATTATCATAG 60
|||||
Db 189 ATTAAGATTATCATAG 205

RESULT 31
AA052408
LOCUS
DEFINITION
mb72h01.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone
IMAGE:334993 5' similar to gb:M21389 KERATIN, TYPE II CYTOSKELETAL
5 (HUMAN); mRNA sequence.
ACCESSION AA052408
VERSION AA052408.1 GI:1543263
KEYWORDS
house mouse.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 96)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:216393
Trace considered overall poor quality
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 1.
FEATURES
Source
1..96
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:334993"
/clone_lib="Soares mouse p3NMf19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCTAGTGGAGCGCCGATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Donaldo. RNA was kindly provided by

```

```

BASE COUNT      22 a      26 c      26 g      22 t
ORIGIN
    Query Match      11.9%; Score 16; DB 27; Length 96;
    Best Local Similarity 100.0%; Pred. No. 2.3e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAATGAGTAAGAAGAA 22          29-JAN-1999
    |||||
Db 70 CAATGAGTAAGAAGAA 85

RESULT 32
AI278373
LOCUS      qm63e02.x1 Soares_placenta.8to9weeks.2NbHP8to9w Homo sapiens cDNA
DEFINITION clone IMAGE:1893434 3', mRNA sequence.
ACCESSION  AI278373
VERSION     AI278373
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 173)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (infoimage.llnl.gov) for further information.
            Insert Length: 1154 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 150.
FEATURES             Location/Qualifiers
             1..173
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:1893434"
             /clone_lib="Soares_placenta.8to9weeks.2NbHP8to9w"
             /dev_stage="two placentae: one from 8 weeks and another
             from 9 weeks post conception"
             /lab_host="DH10B (ampicillin resistant)"
             /note="Organ: placenta; Vector: p773b (pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
             strand cDNA was primed with a Not I - oligo(dT) primer [5',
             TGTTACCAATCTGAAGTCGGAGCGCGGATTTTTTTTTTTT 3'],
             double-stranded cDNA was size selected, ligated to Eco RI
             adaptors (pharmacia), digested with Not I and cloned into
             the Not I and Eco RI sites of a modified p773 vector
             (pharmacia). Library constructed by Bento Soares and
             M.Fatima Bonaldo."
BASE COUNT      52 a      32 c      22 g      67 t
ORIGIN
    Query Match      11.9%; Score 16; DB 44; Length 173;
    Best Local Similarity 100.0%; Pred. No. 1.9e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GATTATCATAGACGA 64
    |||||
Db 88 GATTATCATAGACGA 103

RESULT 33
AV011313/c
LOCUS      AV011313 195 bp mRNA EST 25-AUG-1999

DEFINITION AV011313 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
clone 1110031J17, mRNA sequence.
ACCESSION  AV011313
VERSION     AV011313.1 GI:4788300
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 195)
            Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
            RIKEN Mouse ESTs
            Unpublished (1999)
            On Jun 5, 1998 this sequence version replaced gi:3188818.
JOURNAL    Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@tc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
             1..195
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="1110031J17"
             /clone_lib="Mus musculus 18-day embryo C57BL/6J"
             /sex="mixed"
             /dev_stage="18-day embryo"
BASE COUNT      50 a      50 c      33 g      62 t
ORIGIN
    Query Match      11.9%; Score 16; DB 49; Length 195;
    Best Local Similarity 100.0%; Pred. No. 1.9e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TGAGTAAGAAGAAGA 25
    |||||
Db 44 TGAGTAAGAAGAAGA 29

RESULT 34
AA570404
LOCUS      nk62g08.s1 NCI_CGAP_Schl Homo sapiens cDNA clone IMAGE:1018142 3',
DEFINITION mRNA sequence.
ACCESSION  AA570404
VERSION     AA570404.1 GI:2344384
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 213)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index

```

JOURNAL
COMMENT

Unpublished (1997)
On Nov 29, 1993 this sequence version replaced gi:637706.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1191 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 144.

FEATURES

source

1. .213
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Schl"
/tissue_type="Schwannoma tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Two
pooled bulk Schwannoma tumors. 5' adaptor sequence: 5'
GAATGGCAGCG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
54 a 45 c 29 g 85 t

Query Match 11.9%; Score 16; DB 35; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GATTATCATAGACGA 64

Db 96 GATTATCATAGACGA 111

RESULT 35

A1277304 262 bp mRNA EST 29-JAN-1999
LOCUS qm67f07.x1 Soares.Placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
DEFINITION clone IMAGE:1893829 3', mRNA sequence.
ACCESSION A1277304
VERSION A1277304.1 GI:3899572
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 262)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2287250.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1206 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 241.

FEATURES

source

1. .262
Location/Qualifiers
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone_lib="Soares.Placenta_8to9weeks_2NBHP8to9W"
/dev_stage="two placenta; one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DHI0B (ampicillin resistant)"
/note="Organ: placenta; Vector: p77T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCGCGGATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p77T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaudo."

BASE COUNT 74 a 58 c 37 g 93 t

ORIGIN

Query Match 11.9%; Score 16; DB 44; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GATTATCATAGACGA 64
Db 88 GATTATCATAGACGA 103

RESULT 36

AQ663218 263 bp DNA GSS 23-JUN-1999
LOCUS HS_5463_B2_D01_SP6E RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1039 Col=2 Row=H, genomic survey sequence.
ACCESSION AQ663218.1 GI:5170986
VERSION AQ663218
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 263)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1039 Row: H Column: 2
Seq primer: SP6
Class: BAC ends

FEATURES

source

High quality sequence stop: 263.

Location/Qualifiers

1. .263

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Plate=1039 Col=2 Row=H"

/clone_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Genomic sequence of BAC ends"


```

/strain="N2"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/dev_stage="embryo"
BASE COUNT      84 a      62 c      61 g      91 t      2 others
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 30 ATTAGACAAATAAAC 45

RESULT 40
C51747 300 bp mRNA EST 11-SEP-1997
LOCUS
DEFINITION C51747 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk214b4 3', mRNA sequence.
ACCESSION C51747 GI:2389504
VERSION
KEYWORDS
SOURCE
ORGANISM Caenorhabditis elegans.
REFERENCE
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1400939.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      84 a      63 c      62 g      89 t      2 others
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 29 ATTAGACAAATAAAC 44

RESULT 41
C52581 300 bp mRNA EST 11-SEP-1997
LOCUS
DEFINITION C52581 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk260g2 3', mRNA sequence.

```

```

C52581 300 bp mRNA EST 11-SEP-1997
LOCUS
DEFINITION C52581 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk286b3 3', mRNA sequence.
ACCESSION C52912
VERSION C52912.1 GI:2390669
KEYWORDS EST.
SOURCE
ORGANISM Caenorhabditis elegans.
REFERENCE
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      76 a      64 c      59 g      93 t      8 others
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 25 ATTAGACAAATAAAC 40

RESULT 42
C52912 300 bp mRNA EST 11-SEP-1997
LOCUS
DEFINITION C52912 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk286b3 3', mRNA sequence.
ACCESSION C52912
VERSION C52912.1 GI:2390669
KEYWORDS EST.
SOURCE
ORGANISM Caenorhabditis elegans.
REFERENCE
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      76 a      64 c      59 g      93 t      8 others
ORIGIN

```

```

/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      84 a 63 c 62 g 89 t 2 others
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 29 ATTAGACAAATAAAC 44

RESULT 43
AV178626      300 bp mRNA EST 21-JUL-1999
LOCUS AV178626 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone yk569d5 3', mRNA sequence.
ACCESSION AV178626
VERSION AV178626.1 GI:5558527
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H.,
Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and
Nomoto, H.
Expressed genes in C.elegans
Unpublished (1999)
On May 18, 1998 this sequence version replaced gi:3137286.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
BASE COUNT      84 a 63 c 64 g 89 t
ORIGIN

Query Match      11.9%; Score 16; DB 62; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 28 ATTAGACAAATAAAC 43

RESULT 44
AV184617      300 bp mRNA EST 21-JUL-1999
LOCUS AV184617 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone yk667b11 3', mRNA
sequence.

```

```

ACCESSION AV184617.1 GI:5564518
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H.,
Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and
Nomoto, H.
Expressed genes in C.elegans
Unpublished (1999)
On Mar 16, 1998 this sequence version replaced gi:2961756.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
BASE COUNT      84 a 65 c 62 g 89 t
ORIGIN

Query Match      11.9%; Score 16; DB 62; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
|||||
Db 22 ATTAGACAAATAAAC 37

RESULT 45
AV332351      305 bp mRNA EST 11-NOV-1999
LOCUS AV332351 RIKEN full-length enriched, adult male medulla oblongata
DEFINITION Mus musculus cDNA clone 6330533L04 3', mRNA sequence.
ACCESSION AV332351
VERSION AV332351.1 GI:6372403
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 305)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I.,
Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,
Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, Y.,
Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A.,
Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (1999)
On Dec 20, 1995 this sequence version replaced gi:1133772.
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 21:09:53 ; Search time 9481.22 Seconds
(without alignments)
-63.090 Million cell updates/sec

Title: US-09-090-672B-10
Perfect score: 197
Sequence: 1 ATGGTACCAGTTTCAAATT.....AAGAACTCTATTATGACT 197

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pri.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: gb_vi.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_sy.*
30: em_un.*
31: em_vi.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_hum3.*
39: em_hum4.*
40: gb_pr4.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*

44: gb_htg6.*
45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	197	100.0	177990	43	AC012514	AC012514 Homo sapi
2	146	74.1	163288	45	AC012013	AC012013 Homo sapi
3	24	12.2	108056	8	ATAC005170	AC005170 Arabidops
4	23	11.7	161155	45	AC016910	AC016910 Homo sapi
5	21	10.7	167643	32	CNS01DSZ	AL122057 Homo sapi
6	21	10.7	170846	32	CNS01DTJ	AL132778 Homo sapi
7	20	10.2	72941	11	AC004227	AC004227 Homo sapi
8	20	10.2	77340	43	AC014684	AC014684 Drosophil
9	20	10.2	72941	11	AC004227	AC004227 Homo sapi
10	20	10.2	88071	11	AC002316	AC013680 Arabidops
11	20	10.2	162345	35	AC006415	AC002316 Homo sapi
12	19	9.6	5084	10	HS800872	AC008415 Drosophil
13	19	9.6	74673	40	AC005683	AL110218 Homo sapi
14	19	9.6	100368	11	HS514K20	AC005683 Homo sapi
15	19	9.6	111723	11	HSJ101K10	AL008731 Human DNA
16	19	9.6	116205	41	AC009457	AL080276 Human DNA
17	19	9.6	128247	33	AL133269	AC009457 Drosophil
18	19	9.6	135711	43	AC015596	AL133269 Homo sapi
19	19	9.6	152583	33	HSJ709L21	AL078646 Homo sapi
20	19	9.6	154129	41	AC009427	AC015596 Homo sapi
21	19	9.6	155164	11	AC005165	AC009427 Homo sapi
22	19	9.6	158454	42	AC011107	AC005165 Homo sapi
23	19	9.6	166925	33	AC007027	AC011107 Homo sapi
24	19	9.6	174050	44	AC009863	AC007027 Homo sapi
25	19	9.6	192823	43	AC012149	AC009863 Homo sapi
26	19	9.6	231949	32	HSI099D15	AC012149 Homo sapi
27	19	9.6	237523	1	RPXX04	AL033456 Homo sapi
28	19	9.1	1563	7	BNASLRGA	AJ235273 Rickettsi
29	18	9.1	1563	7	BNESLOGPB	L10737 Brassica na
30	18	9.1	1794	7	BNSL1R	Z21609 B.napus mRN
31	18	9.1	1825	12	MMU13839	Z26914 B.napus (We
32	18	9.1	2139	7	BOSLRG	U13839 Mus musculu
33	18	9.1	2787	34	PFSARP	X52089 Brassica ge
34	18	9.1	2984	7	ATRPCL15G	Z26314 P.falciparu
35	18	9.1	7186	2	AF102623	Z11508 A.thaliana
36	18	9.1	10414	1	MX99740	AF102623 Methanosa
37	18	9.1	26542	34	CEM18	X99740 Mycoplasma
38	18	9.1	38886	34	CELF46H5	Z68507 Caenorhabdi
39	18	9.1	41402	41	AC008437	U41543 Caenorhabdi
40	18	9.1	45617	44	AC016648	AC008437 Homo sapi
41	18	9.1	57784	42	AC012518	AC016648 Homo sapi
42	18	9.1	70390	41	AC010280	AC012518 Homo sapi
43	18	9.1	70952	7	AB024028	AC010280 Homo sapi
44	18	9.1	71293	44	AC016217	AB024028 Arabidops
45	18	9.1	88481	40	AF191071	AC016217 Homo sapi
						AF191071 Homo sapi

ALIGNMENTS

RESULT 1
AC012514
LOCUS AC012514 177990 bp DNA 17-NOV-1999
DEFINITION Homo sapiens clone RP11-837P6, *** SEQUENCING IN PROGRESS ***, 97
unordered pieces.
AC012514
AC012514

AC012514.3	GI:6446645
HTG; HTGS_PHASE1.	
human.	
ORGANISM	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Homidae; Homo.	
1 (bases 1 to 177990)	
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,	
Bodoty,B., Bock,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,	
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,	
David,R., Delgado,O., Deshazo,D., Ding,Y., Doman-Rashid,N.,	
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,	
Forcum-Taney,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,	
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,	
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,	
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,	
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,	
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,	
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,	
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.B., Quiles,M.,	
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,	
Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P.,	
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,	
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,	
Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.	
Direct Submission	
Unpublished	
2 (bases 1 to 177990)	
Worley,K.C.	
Direct Submission	
Submitted (29-Oct-1999) Human Genome Sequencing Center, Department	
of Molecular and Human Genetics, Baylor College of Medicine, One	
Baylor Plaza, Houston, TX 77030, USA	
On Nov 17, 1999 this sequence version replaced gi:6249649.	
* NOTE: This is a "working draft" sequence. It currently	
* consists of 97 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence.	
* as soon as it is available and the accession number will	
* be preserved.	
6897: contig of 6897 bp in length	
6898 6917: gap of unknown length	
6918 13470: contig of 6553 bp in length	
13471 13490: gap of unknown length	
13491 19070: contig of 5580 bp in length	
19071 19090: gap of unknown length	
19091 24387: contig of 5297 bp in length	
24388 24407: gap of unknown length	
24408 29534: contig of 5127 bp in length	
29535 29554: gap of unknown length	
29555 34620: contig of 5066 bp in length	
34621 34640: gap of unknown length	
34641 38332: contig of 3692 bp in length	
38333 38352: gap of unknown length	
38353 42003: contig of 3651 bp in length	
42004 42023: gap of unknown length	
42024 45245: contig of 3222 bp in length	
45246 45265: gap of unknown length	
45266 48418: contig of 3153 bp in length	
48419 48438: gap of unknown length	
48439 51515: contig of 3077 bp in length	
51516 51535: gap of unknown length	
51536 54451: contig of 2916 bp in length	
54452 54471: gap of unknown length	
54472 57383: contig of 2912 bp in length	
57384 57403: gap of unknown length	
57404 60180: contig of 2777 bp in length	
60181 60200: gap of unknown length	
60201 62927: contig of 2727 bp in length	
62928 62947: gap of unknown length	
62948 65653: contig of 2706 bp in length	
65654 65673: gap of unknown length	
65674 68341: contig of 2868 bp in length	
68342 68361: gap of unknown length	
68362 71027: contig of 2866 bp in length	
71028 71047: gap of unknown length	
71048 73713: contig of 2866 bp in length	
73714 73733: gap of unknown length	
73734 76368: contig of 2635 bp in length	
76369 76388: gap of unknown length	
76389 78943: contig of 2555 bp in length	
78944 78953: gap of unknown length	
78954 81471: contig of 2508 bp in length	
81472 81491: gap of unknown length	
81492 83863: contig of 2372 bp in length	
83864 83884: gap of unknown length	
83885 86170: contig of 2287 bp in length	
86171 86190: gap of unknown length	
86191 88472: contig of 2282 bp in length	
88473 88492: gap of unknown length	
88493 90647: contig of 2155 bp in length	
90648 90667: gap of unknown length	
90668 92730: contig of 2063 bp in length	
92731 92750: gap of unknown length	
92751 94742: contig of 1992 bp in length	
94743 94762: gap of unknown length	
94763 96726: contig of 1964 bp in length	
96727 96746: gap of unknown length	
96747 98680: contig of 1934 bp in length	
98681 98700: gap of unknown length	
98701 100627: contig of 1927 bp in length	
100628 100647: gap of unknown length	
100648 102574: contig of 1927 bp in length	
102575 102594: gap of unknown length	
102595 104483: contig of 1889 bp in length	
104484 104503: gap of unknown length	
104504 106367: contig of 1864 bp in length	
106368 106387: gap of unknown length	
106388 108194: contig of 1807 bp in length	
108195 108214: gap of unknown length	
108215 110019: contig of 1805 bp in length	
110020 110039: gap of unknown length	
110040 111832: contig of 1793 bp in length	
111833 111852: gap of unknown length	
111853 113641: contig of 1789 bp in length	
113642 113661: gap of unknown length	
113662 115374: contig of 1713 bp in length	
115375 115394: gap of unknown length	
115395 117089: contig of 1695 bp in length	
117090 117109: gap of unknown length	
117110 118779: contig of 1670 bp in length	
118780 120432: gap of unknown length	
120433 120452: gap of unknown length	
120453 122066: contig of 1614 bp in length	
122067 122086: gap of unknown length	
122087 123382: contig of 1496 bp in length	
123383 123602: gap of unknown length	
123603 125026: contig of 1424 bp in length	
125027 125046: gap of unknown length	
125047 126450: contig of 1414 bp in length	
126451 126480: gap of unknown length	
126481 127880: contig of 1400 bp in length	
127881 127900: gap of unknown length	
127901 129292: contig of 1392 bp in length	
129312 129312: gap of unknown length	
129313 130693: contig of 1381 bp in length	
130713 130694: gap of unknown length	
130714 132078: contig of 1365 bp in length	
132079 132098: gap of unknown length	
132099 133458: contig of 1360 bp in length	
133478 133478: gap of unknown length	
134827 134827: contig of 1349 bp in length	
134847 134847: gap of unknown length	


```

* 153268 153287: gap of unknown length
* 153288 154672: contig of 1385 bp in length
* 154692 154692: gap of unknown length
* 155930 155930: contig of 1258 bp in length
* 155951 155950: gap of unknown length
* 155971 157219: contig of 1249 bp in length
* 157220 157239: gap of unknown length
* 157240 158238: contig of 1059 bp in length
* 158299 158318: gap of unknown length
* 158319 159151: contig of 833 bp in length
* 159152 159171: gap of unknown length
* 159172 160001: contig of 830 bp in length
* 160002 160021: gap of unknown length
* 160022 160828: contig of 807 bp in length
* 160829 160848: gap of unknown length
* 160849 161652: contig of 804 bp in length
* 161653 161672: gap of unknown length
* 161673 162472: contig of 800 bp in length
* 162473 162493: gap of unknown length
* 162493 163288: contig of 796 bp in length.
FEATURES
    source
    1..163288
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-764L14"
BASE COUNT 47826 a 31744 c 33111 g 50313 t 294 others
ORIGIN
Query Match 74.1%; Score 146; DB 45; Length 163288;
Best Local Similarity 99.3%; Pred. No. 4.5e-61;
Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGTTACCCAGTTCAAAATTAACATGGTTATTTTACTTGTTGCCAAATTTAACATTAG 60
Db 122244 ATGTTACCCAGTTCAAAATTAACATGGTTATTTTACTTGTTGCCAAATTTAACATTAG 122185
QY 61 GGAATTTTGGTGGGCTGTATCACTAGAGAAATATATATATATGTCGTGAAGATA 120
Db 122184 GGAATTTTGGTGGGCTGTATCACTAGAGAAATATATATGTCGTGAAGATA 122125
QY 121 ATTTTGAGATAATAGACAGAGAGTTTACATTTCAAGAACAGTTTGGCAGTTGAAG 180
Db 122124 ATTTTGAGATAATAGACAGAGAGTTTACATTTCAAGAACAGTTTGGCAGTTGAAG 122065
QY 181 AATCTATTATATGACT 197
Db 122064 AATCTATTATATGACT 122048
RESULT 3
ATAC005170/c ATAC005170 108056 bp DNA PLN 13-OCT-1998
LOCUS Arabidopsis thaliana chromosome II BAC T29E15 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC005170
VERSION AC005170.1 GI:3738313
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi.
REFERENCE 1 (bases 1 to 108056)
AUTHORS Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
Shen,M., Rensing,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
TITLE Arabidopsis thaliana chromosome II BAC T29E15 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108056)
AUTHORS Rounsley,S.D. and Lin,X.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) The Institute for Genomic Research, 9712

```

```

REFERENCE
AUTHORS Rounsley,S.D.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) The Institute for Genomic Research, 9712
COMMENT Medical Center Dr., Rockville, MD 20850, USA
On Oct 13, 1998 this sequence version replaced gi:3608124.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA
e-mail: rounsley@tigr.org
BAC clone T29E15 is from Arabidopsis chromosome II and is contained
in the YAC clone C108C07.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tldb/at/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.
FEATURES
    source
    1..108056
    /organism="Arabidopsis thaliana"
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /chromosome="II"
    /map="C108C07"
    /clone="T29E15"
    /complement(<1..219)
    /gene="T29E15.1"
    /complement(<1..219)
    /gene="T29E15.1"
    /note="T29E15.1"
    /note="F27L4.1"
    /complement(1..3525)
    /note="overlap with BAC clone F27L4 (AC004482.1..3525).".
    /complement(<1..210)
    /gene="T29E15.1"
    /note="unknown protein, 3' partial"
    /codon_start=1
    /protein_id="AAC63855.1"
    /db_xref="GI:3738314"
    /translation="MFLLILLVFCITVFVFWTNKGAGAEIGKGYKEYKLGDYSTMWL
    QKRVENGKNWKNKIRSLVESKVSGL"
    270..445
    /note="exon predicted by xgrail, quality
    excellent_shadowexon"
    /complement(688..704)
    /note="exon predicted by xgrail, quality marginal"
    /rpt_family="(TAAA)n"
    /complement(774..884)
    /rpt_family="(TAAA)n"
    /complement(777..848)
    /rpt_family="(TAAA)n"
    1007..1064
    /rpt_family="AT-rich"

```

Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 108056)

Rounsley,S.D.
Direct Submission
Submitted (13-OCT-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Oct 13, 1998 this sequence version replaced gi:3608124.
Address all correspondence to:

Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA

e-mail: rounsley@tigr.org
BAC clone T29E15 is from Arabidopsis chromosome II and is contained
in the YAC clone C108C07.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), and NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tldb/at/at.html).

Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.

Location/Qualifiers
1..108056
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="C108C07"
/clone="T29E15"
/complement(<1..219)
/gene="T29E15.1"
/complement(<1..219)
/gene="T29E15.1"
/note="T29E15.1"
/note="F27L4.1"
/complement(1..3525)
/note="overlap with BAC clone F27L4 (AC004482.1..3525)."
/complement(<1..210)
/gene="T29E15.1"
/note="unknown protein, 3' partial"
/codon_start=1
/protein_id="AAC63855.1"
/db_xref="GI:3738314"
/translation="MFLLILLVFCITVFVFWTNKGAGAEIGKGYKEYKLGDYSTMWL
QKRVENGKNWKNKIRSLVESKVSGL"
270..445
/note="exon predicted by xgrail, quality
excellent_shadowexon"
/complement(688..704)
/note="exon predicted by xgrail, quality marginal"
/rpt_family="(TAAA)n"
/complement(774..884)
/rpt_family="(TAAA)n"
/complement(777..848)
/rpt_family="(TAAA)n"
1007..1064
/rpt_family="AT-rich"

```

repeat_region      1103..1166
/rpt_family="AT-rich"
repeat_region      complement(1172..1181)
/rpt_family="(CAA)n"
repeat_region      1959..2114
/rpt_family="AT-rich"
repeat_region      2396..2422
/rpt_family="AT-rich"
/rpt_family="AT-rich"
misc_feature       complement(2741..2858)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
misc_feature       3114..3145
/note="exon predicted by xgrail, quality marginal"
repeat_region      3586..3691
/rpt_family="(CAAAA)n"
mRNA               join(<3687..3956,4809..4933,5018..5055,5136..5246,
5359..5475,5694..5993)
/gene="T29E15.2"
<3687..5993
/gene="T29E15.2"
CDS                join(3687..3956,4809..4933,5018..5055,5136..5246,
5359..5475,5694..5764)
/gene="T29E15.2"
/codon_start=1
/note="unknown protein"
/protein_id="AAC63656.1"
/db_xref="GI:3738315"
/translation="MKNVRFVCKSLVFLVPSFTSRSHIRFTVSAAGASSPNRAIH
CMASDSQSGDGVSSPPNVAAYPSSSSSSASSALDFSLCTRLKTPPRAGWKRDV
KDPESIAHMRIGMLMALISSDIPGVNRKCMKMAIVHDIAEIVGDITPSCGISKEE
KNRRESALEHMKLGGGERAKIEAELRWREYENSPEAKVVRDFDVELILQALEY
EQGFQNIKNAASEIVSRKQK"
complement(3861..3919)
/rpt_family="(GAA)n"
repeat_region      6423..6450
/rpt_family="AT-rich"
mRNA               join(<6738..6789,6916..6986,7073..7154,7261..7376,
7475..7603)
/gene="T29E15.3"
<6738..7603
/gene="T29E15.3"
CDS                join(6738..6789,6916..6986,7073..7154,7261..7376,
7475..7603)
/gene="T29E15.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63657.1"
/db_xref="GI:3738316"
/translation="MSNNELLEIEPMYLOFPPELKMKSCSLYLNKIDNNVAFKVKI
TNRNVCVRPNYGVILPKSTCKVLTMQAKVPSDMQSFERFMIOSLVLAGPGVTAK
VTREMFSGSHVVEETKLRTVVCSTTTNITSPRTGRGFQFC"
complement(7640..7728)
misc_feature       /note="exon predicted by xgrail, quality
excellent_shadowexon"
misc_feature       7643..7711
/note="exon predicted by xgrail, quality excellent"
misc_feature       7881..7943
/note="exon predicted by xgrail, quality marginal"
misc_feature       8342..8432
/note="exon predicted by xgrail, quality excellent"
repeat_region      complement(8539..8578)
/rpt_family="(TAAA)n"
misc_feature       8791..8854
/note="exon predicted by xgrail, quality marginal"
misc_feature       complement(8817..8852)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
misc_feature       complement(8993..9066)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
misc_feature       9139..9226
/note="exon predicted by xgrail, quality marginal"
misc_feature       complement(9146..9242)

```

```

/note="exon predicted by xgrail, quality
marginal_shadowexon"
10777..10792
/note="exon predicted by xgrail, quality good"
complement(11093..11148)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
12104..12164
/note="exon predicted by xgrail, quality excellent"
complement(12228..12321)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(12754..12788)
/note="exon predicted by xgrail, quality excellent"
12885..12931
/note="exon predicted by xgrail, quality good_shadowexon"
complement(12941..13003)
/note="exon predicted by xgrail, quality marginal"
13008..13124
/rpt_family="(TAAAA)n"
repeat_region      13060..13125
/rpt_family="(CAAAA)n"
repeat_region      13258..13343
/rpt_family="(TAAAA)n"
repeat_region      13429..13533
/rpt_family="AT-rich"
repeat_region      complement(13851..14008)
/rpt_family="AT-rich"
15830..15853
/note="exon predicted by xgrail, quality marginal"
15903..15991
/note="exon predicted by xgrail, quality excellent"
complement(16156..16274)
Query Match      12.2% Score 24; DB 8; Length 108056;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 CAAGACAAGTTGGCAGTTGAAG 180
|||||
Db 20343 CAAGACAAGTTGGCAGTTGAAG 20320

RESULT 4
AC016910/c AC016910 161155 bp DNA HTG 08-DEC-1999
LOCUS Homo sapiens clone RP11-460B1, *** SEQUENCING IN PROGRESS ***, 21
DEFINITION unordered pieces.
ACCESSION AC016910
VERSION AC016910.1 GI:6539396
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161155)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 161155)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* Center project name: H.NH0450B01.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```


Contig 20 : length 1709 bp
 Contig 26 : length 12108 bp
 Contig 30 : length 32339 bp
 Contig 29 : length 62473 bp
 Contig 23 : length 2778 bp
 Contig 27 : length 20623 bp
 Contig 22 : length 1711 bp
 Contig 28 : length 23487 bp
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source
 Location/Qualifiers
 1. .170846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone_lib="Caltech-D"
 /clone="C-2568P8"
 367. .484
 /standard_name="AA034525"
 /note="matching EMBL:AA034525; Identified using the e-PCR
 software (G. Schuler)"
 12305. .12491
 /standard_name="AA599294"
 /note="matching EMBL:AA599294; Identified using the e-PCR
 software (G. Schuler)"
 BASE COUNT 50617 a 30842 c 30677 g 49689 t 9021 others
 ORIGIN

Query Match 10.7%; Score 21; DB 32; Length 170846;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 129 ATAAATAGACAGACAGTTTA 149

|||||
 Db 169459 ATAAATAGACAGACAGTTTA 169479

RESULT 7

LOCUS AC014684/c
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 AC014684
 AC014684.1 GI:6436651
 HTG; HTGS_PHASE2.
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 Adams,M. and Venter,J.C.
 Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

Direct Submission
 This sequence was identified as CDM:10210542 by the submitter.
 For further information on this sequence e-mail to fly@celera.com.

COMMENT
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source
 Location/Qualifiers
 1. .55307
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 16942 a 11043 c 11010 g 16312 t
 ORIGIN

Query Match

10.2%; Score 20; DB 43; Length 55307;

Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 AAAAATATATATGGTGC 112

|||||
 Db 26274 AAAAATATATATGGTGC 26255

RESULT 8

LOCUS AC004227 72941 bp DNA PRI 26-FEB-1998
 DEFINITION Homo sapiens chromosome 5, P1 clone 356a8 (LBNL H32), complete
 sequence.
 AC004227 L81382 L81382 L77059 L77058 L81381 L81629 L81804 AC001492
 AC000982 AC000985 AC000983 AC001491 AC000984
 VERSION AC004227.1 GI:2911718
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 72941)
 REFERENCE Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,
 Miguel,T., Lewis,K.D., Fridlyand,J., Alcivare,D., Benke,J.A.,
 Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,
 Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,
 Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,
 Abrasiano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,
 Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.

Sequencing of human chromosome 5
 Unpublished
 2 (bases 1 to 72941)
 REFERENCE Ricke,D.O. and Wagner,R.P.
 Large Scale Sequence Analysis and Annotation with the Sequence
 Comparison Analysis (SCAN) System
 Unpublished
 3 (bases 1 to 72941)
 REFERENCE Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,
 Miguel,T., Lewis,K.D., Fridlyand,J., Alcivare,D., Benke,J.A.,
 Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,
 Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,
 Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,
 Abrasiano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,
 Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.

Direct Submission
 Submitted (26-FEB-1998) Human Genome Center, DOE Joint Genome
 Institute, Lawrence Berkeley National Laboratory, MS 74-157,
 Berkeley, CA 94720, U.S.A.
 Sequence submitted by:
 DOE Joint Genome Institute

COMMENT
 For further information about this sequence, including its location
 and relationship to other sequences, please visit the sequence
 archive website (<http://www-hgc.lbl.gov/sequence-archive.html>) or
 send email to human@genome.lbl.gov

FEATURES
 Location/Qualifiers
 1. .72941
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="5q"
 /clone="356a8"
 /chromosome="5"
 /note="LBNL H32"
 complement(1294. .1824)
 /rpt_family="Alu"
 complement(2001. .2178)
 /note="88% identity mouse vo51c02.r1"
 /db_xref="dbEST:AA611287"
 complement(<2020. .>50956)
 /gene="KIAA0011B"
 complement(join(<2020. .2177,8946. .9186,18985. .19087,

```
22015. .22218.24986. .25093.27876. .28040.34059. .34211,
34797. .34955.37504. .37640.39747. .39924.50823. .>50956))
/ gene="KIA001LB"
/codon_start=3
/evidence_cenot_experimental
/product="KIA001LB"
/protein_id="AAC04379.1"
/db_xref="GI:2911719"
/translation="LPADLTKMHLTENPHQVTHVSSQSGSIASDSGSSLSDIYQ
ATESEVGDVLTRLRPEGVDSDEDEDEIDRTDPLQGRDLVRECLKEPADKTD
VEIQLEFMHOLPAFAVMTMSVRKELCSVMIFVVEOAGAIILEDGOELDSVYILNGT
VEISHPGKVENFPMGNSGEGITTLQKQVMHGIVRTKVDCCOFVCAOODVWRLNHV
EKNTKHYEEGEIVMVEHRELDORSGRKHIVIKATPERLIMHLEHSIVDTYIE
DFLLTYRTEFSLPDVGILKULEFKIDSURDKVTRIVLWVNNHFNDEGPFAMTRFL
BEFKNEDTMNGHLLNIAACAARQVVLQKASRESPLQFSLNGSEKGFIF
VEGVESKADSGLRKGDQIMEVNGSONFENITFMKAVEILRNTHLALTVKTNFKR
LLFTQEKSGVPHPIAKKNSHNSIOHVFDIGTQSKGSKVKANTVSGRKN
IRKILDKTRFSILPPKLFSDGGLSQSDDSIVGTRHCHSLAIMPITGLTSSSPDLL
OPTSMILDFSNPS"
misc_feature
2580. .2637
/ note="GRAIL 2 good exon, frame 1"
repeat_region
3601. .3903
/ rpt_family="Alu"
repeat_region
5735. .5889
/ rpt_family="Alu"
repeat_region
complement(5874. .6290)
/ rpt_family="Alu"
misc_feature
6487. .6899
/ note="97% identity yfilbil.sl"
/ db_xref="dbEST:R06850"
complement(7362. .7859)
/ gene="KIA001LB"
misc_feature
/ note="95% identity yfilbil.rl"
/ db_xref="dbEST:R0689"
complement(8946. .9164)
/ gene="KIA001LB"
misc_feature
/ note="GRAIL 2 good exon, frame 1"
complement(8959. .9186)
/ gene="KIA001LB"
repeat_region
/ note="61% identity KIAA0313"
complement(9570. .9793)
/ rpt_family="MER8"
repeat_region
complement(9926. .10218)
/ rpt_family="AluSx"
repeat_region
complement(10364. .10657)
/ rpt_family="Alu"
repeat_region
10658. .10963
/ rpt_family="Alu"
repeat_region
complement(11742. .12035)
/ rpt_family="Alu"
repeat_region
12493. .12800
/ rpt_family="Alu"
repeat_region
13239. .13547
/ rpt_family="Alu"
repeat_region
13656. .13678
/ note="(T)23"
/ rpt_type=tandem
/ rpt_unit=r
repeat_region
complement(13658. .14011)
/ rpt_family="AluSc"
repeat_region
complement(14396. .14675)
/ rpt_family="Alu"
repeat_region
complement(16140. .16255)
/ rpt_family="Alu"
repeat_region
16160. .16180
/ note="(T)21"
/ rpt_type=tandem
/ rpt_unit=r
repeat_region
complement(16458. .17179)
/ rpt_family="LTR12"
misc_feature
complement(18979. .19084)
/ gene="KIA001LB"
/ note="GRAIL 2 excellent exon, frame 1"
```

```
misc_feature
complement(18985. .19086)
/ gene="KIA001LB"
/ note="82% identity KIAA0313"
complement(20614. .20759)
/ rpt_family="MER20"
repeat_region
20960. .21305
/ rpt_family="MER1B"
misc_feature
complement(22015. .22218)
/ gene="KIA001LB"
/ note="GRAIL 2 excellent exon, frame 2"
complement(22020. .22220)
/ gene="KIA001LB"
/ note="74% identity KIAA0313"
22476. .22511
/ note="(ATT)9"
/ rpt_type=tandem
/ rpt_unit=ATT
repeat_region
complement(22488. .22823)
/ rpt_family="Alu"
misc_feature
23112. .23397
/ note="99% identity aa48d12.sl"
/ db_xref="dbEST:AA490888"
24394. .24751
/ rpt_family="Alu"
misc_feature
complement(24986. .25093)
/ gene="KIA001LB"
/ note="GRAIL 2 good exon, frame 1"
complement(24995. .25096)
/ gene="KIA001LB"
misc_feature
/ note="94% identity KIAA0313"
25480. .25632
/ rpt_family="Alu"
repeat_region
complement(25781. .27005)
/ rpt_family="L1"
misc_feature
complement(27877. .28035)
/ gene="KIA001LB"
/ note="81% identity KIAA0313"
complement(28688. .28900)
/ rpt_family="L1"
repeat_region
28703. .29059
/ rpt_family="AluSp"
repeat_region
29015. .29036
/ note="(A)22"
/ rpt_type=tandem
/ rpt_unit=A
repeat_region
complement(29325. .29653)
/ rpt_family="Alu"
repeat_region
30392. .30799
/ rpt_family="Alu"
repeat_region
31178. .32802
/ rpt_family="MER42"
repeat_region
complement(32183. .32556)
/ rpt_family="AluSx"
repeat_region
33615. .33964
/ rpt_family="Alu"
misc_feature
complement(34058. .34216)
/ gene="KIA001LB"
/ note="88% identity KIAA0313"
complement(34059. .34211)
/ gene="KIA001LB"
misc_feature
/ note="GRAIL 2 excellent exon, frame 0"
complement(34797. .34955)
/ gene="KIA001LB"
/ note="GRAIL 2 excellent exon, frame 0"
```

Query Match 10.2%; Score 20; DB 11; Length 72941;
Best Local Similarity 100.0%; Pred.No. 4.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 GAAAAATATATATTGGTG 111

|||||

Db 44348 GAAAAATATATATTGGTG 44367

	Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W., Fasman,K.H., Lander,E.S., McKernan,K., Munro,C.C., Richardson,P., Barna,N., Chang,A., Cooke,P., Daly,M.J., Devon,K., Dewar,K., Forrest,C., Gage,D., Geraghty,K., Guittau,G., Hagos,B., Huang,J., Jacotot,L., Lane,M., Lee,K., Mackenzie,J., Marquis,N., McDermott,J., Molla,M., Moloney,N., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Olotu,A., Peterson,K., Rollins,G., Spencer,J., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Traish,A., Wilmer,F., Zemtseva,I. and Zody,M.
AUTHORS	
TITLE	Direct Submission
JOURNAL	Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 88071)
AUTHORS	Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K., DuRette,B., Etemadi,S., Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S., Geraghty,K., Gilmarin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L., MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam.R., Soohoo,S., Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo.A., Wagner,A., Wheeler,J., Wu,Y., Ye.W.J., Zemtseva,I., Zhao,J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (03-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	4 (bases 1 to 88071)
AUTHORS	Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K., DuRette,B., Etemadi,S., Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S., Geraghty,K., Gilmarin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L., MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam.R., Soohoo,S., Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo.A., Wagner,A., Wheeler,J., Wu,Y., Ye.W.J., Zemtseva,I., Zhao,J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (05-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	5 (bases 1 to 88071)
AUTHORS	Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geary,K., Grant,G., Hages,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stoianovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A., Wheeler,J., Wu.Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (08-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker. Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html. Location/Qualifiers
FEATURES	
RESULT	9
AC012680/c	
LOCUS	Arabidopsis thaliana chromosome I clone TAMU-Tll111, *** SEQUENCING IN PROGRESS ***, 1 ordered pieces.
DEFINITION	Arabidopsis thaliana
ACCESSION	AC012680
VERSION	GI:6325535
KEYWORDS	HTG; HTGS-PHASE2
SOURCE	thalia cress.
ORGANISM	Eukaryota: Viridiplantae: Streptophyta: Tracheophyta: eudicotyledons: Spermatophyta: Magnoliophyta: euclidyedons; core eucots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsies.
REFERENCE	1 (bases 1 to 77340)
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M., Koo.H., Fujii,C.Y., Uterback.T., Barnstead,M.E., Bowman,C.L., White,O., Neriman,W.C. and Fraser,C.M.
TITLE	Arabidopsis thaliana 'TAMU' BAC 'Tll111' genomic sequence near marker agp64
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 77340)
AUTHORS	Lin.X. and Kaul.S.
TITLE	Direct Submission
JOURNAL	Submitted (03-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA, xlinet@ig.org
COMMENT	* On Nov 10, 1999 this sequence version replaced gi:5223631. * NOTE: This is a 'working draft' sequence. It currently * consists of 1 contigs. Gaps between the contigs * are represented as runs of N. The order of the pieces * is believed to be correct as given, however the sizes * of the gaps between them are based on estimates that have * provided by the submitor * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved * * 1 77340: contig of 77340 bp in length.
BASE COUNT	24601 a 14111 c 13706 g 24916 t
ORIGIN	Location/Qualifiers
source	/organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="I" /clone="TAMU-Tll111"
Query Match	10.2%; Score 20; DB 43; Length 77340;
Best Local Similarity	100.0%; Pred. No. 4.3;
Matches	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	84 TATCACTAGAAAAAATATATA 103
Dd	5574 TATCACTAGAAAAAATATATA 5555
RESULT	10
AC002316	
LOCUS	Arabidopsis thaliana
DEFINITION	Homo sapiens chromosome 17, clone HCIT11023, complete sequence.
ACCESION	AC002316
VERSION	GI:3406048
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:	
Eutheria: Primates: Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 88071)
AUTHORS	Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens chromosome 17, clone HCIT11023
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 88071)

```
source      1. 88071
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HC1111023"
            /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
            1-194)"
            /map="17"
            /chromosome="17"
            complement(236..380)
            /rpt_family="MIR"
            382..679
            /rpt_family="AluSg"
            680..701
            /rpt_family="(TAAA)n"
            complement(706..796)
            /rpt_family="LMB4"
            complement(953..11279)
            /rpt_family="AluSg"
            complement(1335..1523)
            /rpt_family="AluSg/x"
            1531..1762
            /rpt_family="AluJo"
            1766..2198
            /rpt_family="MER65a"
            complement(2223..2515)
            /rpt_family="AluSx"
            complement(2522..2784)
            /rpt_family="LIME"
            3202..3513
            /rpt_family="AluSg"
            4041..4339
            /rpt_family="AluSg"
            complement(5016..5226)
            /rpt_family="LMB4"
            5255..5612
            /rpt_family="MLT1A1"
            5639..5940
            /rpt_family="AluSx"
            7258..7555
            /rpt_family="AluSg"
            7558..7844
            /rpt_family="AluSx"
            7846..7874
            /rpt_family="(TAAA)n"
            8127..8358
            /rpt_family="L2"
            complement(9267..9304)
            /rpt_family="MIR"
            9305..9506
            /rpt_family="MIR"
            10123..10242
            /rpt_family="L2"
            complement(10389..10682)
            /rpt_family="AluSg"
            complement(11431..11741)
            /rpt_family="AluSx"
            12502..12822
            /rpt_family="AluSx"
            12823..12981
            /rpt_family="AluSg"
            complement(13045..13551)
            /rpt_family="L2"
            13552..13864
            /rpt_family="AluSx"
            complement(13865..14044)
            /rpt_family="L2"
            14505..14721
            /rpt_family="LIMC5"
            14722..15022
            /rpt_family="AluJb"
            15023..15043
            /rpt_family="LIMC5"
            15044..15350
```

```
repeat_region      /rpt_family="AluSx"
                    15351..15553
                    /rpt_family="LIMC5"
                    complement(15562..15831)
                    /rpt_family="AluJb"
                    15832..16035
                    /rpt_family="LIMC5"
                    complement(16350..16639)
                    /rpt_family="AluSx"
                    18344..18829
                    /rpt_family="AluJb8"
                    18830..18852
                    /rpt_family="(TAAA)n"
                    20480..20605
                    /rpt_family="(CCCCG)n"
                    20661..20703
                    /rpt_family="GC-rich"
                    20720..20824
                    /rpt_family="GC-rich"
                    complement(21825..21867)
                    /rpt_family="A-rich"
                    22587..22716
                    /rpt_family="AluJb"
                    complement(23096..23192)
                    /rpt_family="L2"
                    complement(23372..23677)
                    /rpt_family="AluSg"
                    complement(23687..23990)
                    /rpt_family="AluY"
                    24158..24457
                    /rpt_family="AluY"
                    25844..25958
                    /rpt_family="AluSx"
                    complement(26157..26234)
                    /rpt_family="L2"
                    27152..27453
                    /rpt_family="AluSg"
                    27824..27821
                    /rpt_family="MIR"
                    27860..28029
                    repeat_region
```

Query Match 10.2% Score 20; DB 11: Length 88071;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TAGAAAAATATATATATGG 109
|||||
Db 12121 TAGAAAAATATATATATGG 12140

RESULT 11
AC006415/c
LOCUS AC006415.162345 bp DNA INV 09-JUL-1999
DEFINITION Drosophila melanogaster, chromosome 2L, region 40B-40C, BAC clone
BACR22A22, complete sequence.
ACCESSION AC006415
VERSION AC006415.11 GI:5430741
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 162345)
AUTHORS Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Ling,H.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Richards,S., Sequeira,A., Sethi,H.,
Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M.,
Zieran,L.L. and Rubin,G.M.

TITLE JOURNAL REFERENCE AUTHORS	Sequencing of Drosophila chromosome 2L, region 40B-40C Unpublished (1998) 2 (bases 1 to 162345) Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhorff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE JOURNAL COMMENT	Direct Submission Submitted (19-JAN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Jul 9, 1999 this sequence version replaced gi:5209360. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720
FEATURES source	For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgpf@fruitfly.berkeley.edu . Location/Qualifiers 1..162345 /organism="Drosophila melanogaster" /strain="y: cn bw sp" /db_xref="taxon:7227" /chromosome="2L" /map="40B-40C" /clone_lib="RPCR-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)" /clone="BAC BACR22A22 (D533)" BASE COUNT 48827 a 32587 c 32005 g 48926 t ORIGIN Query Match 10.2%; Score 20; DB 35; Length 162345; Best Local Similarity 100.0%; Pred. No. 3.9; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Qy 93 AAAATATATATATTGGTGC 112 Db 136676 AAAATATATATATTGGTGC 136657 RESULT 12 HSM800872 LOCUS DEFINITION Homo sapiens mRNA; CDNA DKF2p434A163 (from clone DKF2p434A163). ACCESSION AL110218 VERSION AL110218.1 GI:5817150 KEYWORDS SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 5084) Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S. TITLE Direct Submission JOURNAL Submitted (13-AUG-1999) MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); sequenced by AGOWA within the cDNA sequencing consortium of the German Genome Project. This clone is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de . Location/Qualifiers 1..5084 /organism="Homo sapiens"

/db_xref="taxon:9606" /clone="DKF2p434A163" /clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="adult" /tissue_type="testis" /notes="corresponding STS: EMBL:HS09965" 1..4964 /gene="DKF2p434A163" 1..4964 /gene="DKF2p434A163" /note="strong similarity to C. elegans T27B1.1" /codon_start=3 /product="hypothetical protein" /protein_id="CAB53678.1" /db_xref="GI:5817151" /translation="EDWGDKVICNKFIOTSVCLOWPAEYIIVFGLAEGKYRLANT KTKSSTIYGESYVVSCTNCSGKGLSHGADGTVIRFFDEGSGESQGLVNHPC PPYALAWATNSIVAAGCDRKIVAIGKEGHMLQTFDYSRDPQEREFTVSSPGQSVV LGSYDLRFNFWIPRKSINWEAKPEKITNLTYITLAWKRDGSRCLGTCGGQEQVD CCLRRSYKNKFELTYVPSQIVKNSLSTGRVVLKSHVGYEVEVKILGKERYLVAH TSTLLGLDNTNRLSELIAWQSGGNEKYFFENENVMIFNAGELTYVEGNDTIGS VRTEFMPHLISVIRNERCQRTDNKKLAYLIDIKTIAIVDLIGSGYNTGVSHESRV DWLENDTGKLLFRDKRLRLHYDIESCKTMLNFCSTMQWPQSDVLVAQNRSL CVYNIEAPERTMTFTIRGDVIGLERGGKTEVMVEGTVTAITDEGLIEFTGTAID DGNIRATAFLETMTPEETPEANMKTLKLAEROLHIAERCFALGOVAKARFLHE TNEIADQVSREYEGEGTDFYQVRARLAMEKNYKLAEMIFLEQNAVEAMGYOELHR WDECIAAVEAGHPALEKLRSSYOWIMDTQOERAGELQESOGDGLAAISLVKLAGL PAKAARLVLTRELLANTELVELHITAAIKGELYERAGDLFEKIHHPQKALECYRGN AFMKAVELARLAFPEVYVYKLEAWGDHLVQOKQDAINHIYIARCSIKAEALGAR OWKKAIYILDQFNNTASKYYPVLAQHYASLOEYEAELYTKGDKRTKDAIDMYTAG RWQAHKIAMKMRPEDVSVLYITQAEKMEKQGYEARLYVTVQEPDLAITMYKHH KLVDMDIRLVGKHHPDLLSDTHLHGLKELEAGRIQEAHVYLEAOEWKATVMYRAS GLWEEAYRVARTOGGANAKHVAYLWAKSLGGEAAYVRLNKLGLLEAVDHAADNCSF EFAPFELSRALKHKTPEVHLKYAMFLEDEGKFEEAEAEFRACKPEAVLMFVHQDM EAAQVAAEHDPDSVAELVVGQARGALEEKDFQKAEGLLLRAQRPLALNYKEAGLV SDALRICKDYPSQLEALOEYERATKKGARGVEFVQARHWEQAGEYSRAVDCYL KVRDSNGSLAEKCKWMAELSIFLPPQRNMEVVLAVGQLIGIKHSAEAELYNL DLVKEADAFIEGEEWNKAKRVAKELDPYEDVDQHYEFLKNOCKHSDLSGVGVIA ALDLYVEQGDWDCIETATKQNYKILHKYVALYATHIREGSSAQLALVVOHGAPAN PQPNYIKRIFTDMVSSPGTCAEAYHSHADLRDLVFNLCENLVKSEANSAPAEFEK TMLLIHHYATRSAAQSVKQLETVAAKLSVLSLRLTQLLPVDRKAFYGAAPAAVGM NMAFIFLRFELDLTDAIEGTLGDLSDFDIDPEVFLPAKOHVGAPEAREVEVDW VLTVMDQRLQVLPDRGAYEASLVAASTGVRLPCLITGYTPILRNKTEFRRPQKA ANKDNWNKFLMAIKTSHSPVCQDVLKFIQWCGGLPSTSPSQ" polyA_signal 5001..5006 polyA_site 5011 BASE COUNT 1401 a 1143 c 1434 g 1106 t ORIGIN Query Match 9.6%; Score 19; DB 10; Length 5084; Best Local Similarity 100.0%; Pred. No. 20; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Qy 152 ATTTACAAGAACAGTTTG 170 Db 675 ATTTACAAGACAGTTTG 693 RESULT 13 AC005683 LOCUS AC005683 74673 bp DNA PRI 05-MAY-1999 DEFINITION Homo sapiens clone RG062N11, complete sequence. ACCESSION AC005683 VERSION AC005683.2 GI:4753225 KEYWORDS HTG. SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 74673) AUTHORS Waterston,R.H.
--

TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 74673)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 74673)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On May 5, 1999 this sequence version replaced gi:4204336.
 FEATURES
 source
 1. .74673
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RG062N11"
 BASE COUNT 23089 a 14472 c 14794 g 22318 t
 ORIGIN
 Query Match 9.68; Score 19; DB 40; Length 74673;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 72 TTGTGGGTCTGTTATCACT 90
 |||||
 Db 6100 TTGTGGGTCTGTTATCACT 6118
 RESULT 14
 HS514K20
 LOCUS
 DEFINITION HS514K20 100368 bp DNA PRI 23-NOV-1999
 Human DNA sequence from PAC 514K20 on chromosome 6p22.3-24.3. EST,
 CA repeats and STS.
 ACCESSION AL008731
 VERSION AL008731.1 GI:2769540
 KEYWORDS 6p22.3-24.3; repeat polymorphism.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 100368)
 AUTHORS Wild,A.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-1998) Chromosome 6 Project Group (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk
 On Jan 13, 1998 this sequence version replaced gi:2598467.
 IMPORTANT: This sequence is the entire insert of clone 514K20.
 During sequence assembly data are compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre chromosome 6 mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6/
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The true left end of clone 514K20 is at 1 in this sequence. The true right end of clone 514K20 is at 100368.
 514K20 is from the library RPC13 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong.
 For further details see http://bacpac.med.buffalo.edu/.
 FEATURES
 source
 1. .100368
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p22.3-24.3"
 /clone="RP3-514K20"
 /clone_lib="RPC1-3"
 239, 534
 repeat_region
 /note="AluJo repeat: matches 1. .294 of consensus"
 1723. .2047
 misc_feature
 /note="match: 267338 STS containing (CA) repeat"
 1855. .1877
 repeat_region
 /note="11 copies of CA 100% conserved; differs from 267338"
 1882. .2174
 repeat_region
 /note="AluJo repeat: matches 2. .301 of consensus"
 3503. .3808
 repeat_region
 /note="AluX repeat: matches 2. .301 of consensus"
 4742. .5056
 repeat_region
 /note="AluJo repeat: matches 1. .301 of consensus"
 5276. .5897
 repeat_region
 /note="L1MB3 repeat: matches 291. .923 of consensus"
 6167. .6299
 repeat_region
 /note="L1MB3 repeat: matches 767. .908 of consensus"
 9707. .9975
 repeat_region
 /note="MLT1B repeat: matches 390. .113 of consensus"
 10357. .10552
 repeat_region
 /note="MLT1C repeat: matches 466. .268 of consensus"
 10571. .11046
 repeat_region
 /note="MER4A2 repeat: matches 1. .505 of consensus"
 11042. .11286
 repeat_region
 /note="MLT1C repeat: matches 247. .1 of consensus"
 12252. .12370
 repeat_region
 /note="MLT1B repeat: matches 5. .120 of consensus"
 14219. .14506
 misc_feature
 /note="match: 216457 STS containing (CA) repeat"
 14279. .14319
 repeat_region
 /note="20 copies of CA 100 % conserved; differs from 216457; D6S359"
 15184. .15235
 repeat_region
 /note="2 copies of 26 mer 98 % conserved"
 18638. .18929
 repeat_region
 /note="AluX repeat: matches 302. .6 of consensus"
 19749. .20135
 repeat_region
 /note="MLT1C repeat: matches 23. .466 of consensus"
 21690. .21761
 repeat_region
 /note="MIR repeat: matches 186. .256 of consensus"
 23026. .23327
 repeat_region
 /note="AluX repeat: matches 1. .302 of consensus"
 23328. .23478
 repeat_region
 /note="AluJb repeat: matches 122. .273 of consensus; incomplete repeat"
 24403. .24786
 repeat_region
 /note="L1MD2 repeat: matches 469. .884 of consensus"
 24779. .24937
 repeat_region
 /note="MER42B repeat: matches 1149. .1300 of consensus"
 26000. .26282
 repeat_region
 /note="AluJb repeat: matches 293. .1 of consensus"
 29619. .29921
 repeat_region
 /note="AluY repeat: matches 2. .301 of consensus"
 31380. .31670
 repeat_region
 /note="AluSq repeat: matches 303. .1 of consensus"
 31687. .31753
 repeat_region
 /note="MR2 repeat: matches 141. .75 of consensus"
 31926. .32455
 repeat_region
 /note="L1MA8 repeat: matches 1035. .514 of consensus"
 32475. .32737
 repeat_region
 /note="L1MB1 repeat: matches 566. .302 of consensus"
 32859. .33179
 repeat_region
 /note="AluJo repeat: matches 302. .1 of consensus"

repeat_region 33831..33913 /note="MLTIC repeat: matches 89. .1 of consensus"
repeat_region 34014..34307 /note="AluY repeat: matches 1. .298 of consensus"
repeat_region 35042..35117 /note="MIR2 repeat: matches 146. .74 of consensus"
repeat_region 35186..35487 /note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 36019..36447 /note="MLTIC repeat: matches 456. .1 of consensus"
repeat_region 38175..38477 /note="AluXp repeat: matches 302. .1 of consensus"
repeat_region 38520..38643 /note="MIR2 repeat: matches 145. .16 of consensus"
repeat_region 39015..39337 /note="L1ME1 repeat: matches 432. .661 of consensus"
repeat_region 39457..39631 /note="AluJo repeat: matches 125. .299 of consensus;
incomplete repeat"
repeat_region 39650..39741 /note="L1MA2 repeat: matches 964. .1055 of consensus"
repeat_region 40095..40319 /note="L1PA14 repeat: matches 572. .896 of consensus"
repeat_region 41998..42078 /note="MIR repeat: matches 144. .56 of consensus"
repeat_region 42369..42595 /note="MER28 repeat: matches 1. .230 of consensus"
repeat_region 42654..42911 /note="AluSp repeat: matches 258. .1 of consensus;
incomplete repeat"
repeat_region 42914..43132 /note="TIGGER2 repeat: matches 2472. .2708 of consensus"
repeat_region 43331..43631 /note="AluJo repeat: matches 1. .300 of consensus"
repeat_region 43633..43673 /note="MERSA repeat: matches 69. .109 of consensus"
repeat_region 43866..44166 /note="AluXs repeat: matches 1. .301 of consensus"
repeat_region 45157..45192 /note="18 copies of 2 mer 89 % conserved"
repeat_region 45610..46915 /note="AluXs repeat: matches 302. .1 of consensus"
repeat_region 47459..47740 /note="AluXs repeat: matches 15. .297 of consensus"
repeat_region 48084..48229 /note="MLT1A2 repeat: matches 1. .149 of consensus"
repeat_region 48230..48585 /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region 48586..48820 /note="MLT1B repeat: matches 154. .374 of consensus"
repeat_region 49766..49791 /note="13 copies of 2 mer 100 % conserved"
repeat_region 49951..49982 /note="16 copies of 2 mer 88 % conserved"
repeat_region 50250..50551 /note="AluSp repeat: matches 1. .303 of consensus"
repeat_region 51359..51922 /note="L1ME3A repeat: matches 630. .58 of consensus"
repeat_region 51805..51923 /note="L1MA1 repeat: matches 178. .57 of consensus"
repeat_region 52531..52831 /note="AluSp repeat: matches 3. .302 of consensus"
repeat_region 53351..53654 /note="AluSp repeat: matches 1. .303 of consensus"
repeat_region 54460..54843 /note="MST1D repeat: matches 394. .1 of consensus"
repeat_region 55614..55914 /note="AluSp repeat: matches 1. .303 of consensus"
repeat_region 58484..58593 /note="MSTC repeat: matches 1. .110 of consensus"
repeat_region 58597..58702 /note="MSTC repeat: matches 299. .405 of consensus"
repeat_region 59438..59525

/note="L1MB7 repeat: matches 298. .386 of consensus"
61103..61222 /note="AluY repeat: matches 1. .124 of consensus;
incomplete repeat"
repeat_region 61225..62555 /note="PR5 repeat: matches 2438. .807 of consensus"
repeat_region 62556..62807 /note="AluY repeat: matches 44. .301 of consensus;
incomplete repeat"
repeat_region 62868..63041 /note="MER5A repeat: matches 182. .2 of consensus"
repeat_region 64497..64784 /note="AluB repeat: matches 297. .1 of consensus"
repeat_region 66089..66144 /note="MIR repeat: matches 86. .141 of consensus"
repeat_region 67000..67303
Query Match 9.6%; Score 19; DB 11; Length 100368;
Best Local Similarity 100.0%; Pred.No.13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 128 GATAATTAGACACAGT 146
|||||
Db 82127 GATAATTAGACACAGT 82145
|||||
RESULT 15
LOCUS HSJ101K10 111723 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 101K10 on chromosome 6q25-26,
complete sequence.
ACCESSION AL080276
VERSION AL080276.9 GI:5763753
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 111723)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonesrequest@sanger.ac.uk
COMMENT On Aug 24, 1999 this sequence version replaced gi:5679778.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence is
the entire insert of clone 101K10. This sequence has been finished
according to sequence map criteria as follows. An attempt is made
to resolve all sequencing problems, such as compressions and
repeats, but not necessarily within known annotated human repeat
sequence elements (e.g. Alu). Where the sequence is ambiguous,
there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
101K10 is from the library RPC1-1 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pcYAC2.
FEATURES
Location/Qualifiers
1..111723
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

/chromosome="6"
/map="q25-26"
/clone-lib="RBC1-1"
/clone-RP1-101K10"

BASE COUNT      35439 a 20986 c 21291 g 34007 t
ORIGIN

Query Match          9.6%: Score 19; DB 11; Length 111723;
Best Local Similarity 100.08; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 TTTTGAGATAATTAGACAA 140
      |||||
Db 22529 TTTTGAGATAATTAGACAA 22511

RESULT 16
AC009457/c
LOCUS
DEFINITION
AC009457 116205 bp DNA HTG 29-SEP-1999
Drosophila melanogaster chromosome 3 clone BACR25A06 (D1082)
RPCI-98 25.A.6 map 88E-88E strain y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 106 unordered pieces.
AC009457
VERSION
AC009457.3 GI:5932510
KEYWORDS
HTG: HTGS-PHASE1.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 116205)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chave, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shit, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 116205)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chave, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shit, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 29, 1999 this sequence version replaced gi:5931387.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 106 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 472: contig of 472 bp in length
* 473 552: gap of unknown length
* 553 1038: contig of 486 bp in length

```

```

1118: gap of unknown length
1148: contig of 630 bp in length
1828: gap of unknown length
1749
2432: contig of 604 bp in length
1829
2512: gap of unknown length
2433
3254: contig of 742 bp in length
3255
3334: gap of unknown length
3335
4000: contig of 666 bp in length
3335
4080: gap of unknown length
4001
4838: contig of 738 bp in length
4881
4898: gap of unknown length
4819
5405: contig of 507 bp in length
5406
5485: gap of unknown length
5406
6563: contig of 1078 bp in length
6564
6643: gap of unknown length
6644
7712: contig of 1069 bp in length
7713
7792: gap of unknown length
7793
8542: contig of 750 bp in length
8543
8622: gap of unknown length
8543
9136: contig of 514 bp in length
8543
9216: gap of unknown length
9137
9811: contig of 595 bp in length
9812
9851: gap of unknown length
9812
10285: contig of 374 bp in length
9892
10345: gap of unknown length
10266
11017: contig of 672 bp in length
10346
11018: gap of unknown length
11018
11789: contig of 692 bp in length
11098
11859: gap of unknown length
11790
12484: contig of 615 bp in length
11870
12584: gap of unknown length
12485
13692: contig of 1128 bp in length
12565
13772: gap of unknown length
13693
14746: contig of 974 bp in length
13773
14826: gap of unknown length
14747
15789: contig of 963 bp in length
14827
15869: gap of unknown length
15790
16819: contig of 950 bp in length
15870
16899: gap of unknown length
16820
18094: contig of 1195 bp in length
16900
18174: gap of unknown length
18095
19338: contig of 1164 bp in length
18175
19418: gap of unknown length
19339
20443: contig of 1025 bp in length
19419
20523: gap of unknown length
20444
21557: contig of 1034 bp in length
20524
21637: gap of unknown length
21558
22483: contig of 846 bp in length
21638
22563: gap of unknown length
22484
23643: contig of 1080 bp in length
22564
23743: gap of unknown length
23644
24701: contig of 978 bp in length
23724
24781: gap of unknown length
24702
25914: contig of 1133 bp in length
24782
25994: gap of unknown length
25915
26977: contig of 983 bp in length
25995
27057: gap of unknown length
26978
28157: contig of 1100 bp in length
27058
28237: gap of unknown length
28158
29386: contig of 1149 bp in length
28238
29466: gap of unknown length
29387
30141: contig of 675 bp in length
29467
30221: gap of unknown length
30142
31046: contig of 825 bp in length
30222
31126: gap of unknown length
31047
32575: contig of 1449 bp in length
31127
32655: gap of unknown length
32576
33532: contig of 877 bp in length
32656
33612: gap of unknown length
33533
34152: contig of 540 bp in length
33613
34232: gap of unknown length
34153
35240: contig of 1008 bp in length
34233
35320: gap of unknown length
35241

```

```
*
* 33321 36111: contig of 791 bp in length
* 36112 36191: gap of unknown length
* 37056: contig of 865 bp in length
* 37136: gap of unknown length
* 37137 38947: contig of 1811 bp in length
* 38948 39027: gap of unknown length
* 39028 40211: contig of 1184 bp in length
* 40212 40291: gap of unknown length
* 40292 41628: contig of 1337 bp in length
* 41629 41708: gap of unknown length
* 41709 42827: contig of 1119 bp in length
* 42828 42907: gap of unknown length
* 43008 43329: contig of 1022 bp in length
* 43330 44009: gap of unknown length
* 44010 45197: contig of 1188 bp in length
* 45198 45277: gap of unknown length
* 45278 46489: contig of 1212 bp in length
* 46490 46569: gap of unknown length
* 46570 47667: contig of 1098 bp in length
* 47668 47747: gap of unknown length
* 47748 49123: contig of 1376 bp in length
* 49124 49203: gap of unknown length
* 49204 50931: contig of 1728 bp in length
* 50932 51011: gap of unknown length
* 51012 52120: contig of 1109 bp in length
* 52121 52200: gap of unknown length
* 52201 53648: contig of 1448 bp in length
* 53649 53728: gap of unknown length
* 53729 55170: contig of 1442 bp in length
* 55171 55250: gap of unknown length
* 55251 56886: contig of 1636 bp in length
* 56887 56966: gap of unknown length
* 56967 58781: contig of 1815 bp in length
* 58782 58861: gap of unknown length
* 58862 60422: contig of 1381 bp in length
* 60423 60522: gap of unknown length
* 60523 62554: contig of 2032 bp in length
* 62555 62634: gap of unknown length
* 62635 64871: contig of 2237 bp in length
* 64872 64951: gap of unknown length
* 64952 66791: contig of 1840 bp in length
* 66792 66871: gap of unknown length
* 66872 69289: contig of 2418 bp in length
* 69290 69369: gap of unknown length
* 69370 72789: contig of 3420 bp in length
* 72790 72869: gap of unknown length
* 72870 74329: contig of 1460 bp in length
* 74330 74409: gap of unknown length
* 74410 77754: contig of 3345 bp in length
* 77755 77834: gap of unknown length
* 77835 82129: contig of 4295 bp in length
* 82130 82209: gap of unknown length
* 82210 87148: contig of 4939 bp in length
* 87149 87228: gap of unknown length
* 87229 87849: contig of 621 bp in length
* 87850 87929: gap of unknown length
* 87930 88484: contig of 555 bp in length
* 88485 88564: gap of unknown length
* 88565 89346: gap of unknown length
* 89347 89334: contig of 588 bp in length
* 89335 90014: gap of unknown length
* 90015 90681: contig of 666 bp in length
* 90682 90760: gap of unknown length
* 90761 91355: contig of 595 bp in length
* 91356 91435: gap of unknown length
* 91436 92038: contig of 603 bp in length
* 92039 92118: gap of unknown length
* 92119 92774: contig of 656 bp in length
* 92775 92854: gap of unknown length
* 92855 93544: contig of 590 bp in length
* 93545 93624: gap of unknown length
* 93625 94104: contig of 480 bp in length
*
```

```
Query Match          9.6%; Score 19; DB 41; Length 116205;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 CTAGAAAAATATATATT 107
    |||||
Db 16478 CTAGAAAAATATATATT 16460
```

RESULT 17

```
AL133269          DNA          HTG          03-DEC-1999
LOCUS             Homo sapiens chromosome 6 clone RP3-468K3 map q25-26, ***
DEFINITION        SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION         AL133269
VERSION           AL133269.2 GI:6523580
KEYWORDS          HTG; HTGS_PHASE1.
SOURCE            human.
ORGANISM          Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE         1 (bases 1 to 128247)
AUTHORS           Plumb,B.
TITLE             Direct Submission
JOURNAL           Submitted (03-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
                  requests: clonerequest@sanger.ac.uk
COMMENT           On Dec 3, 1999 this sequence version replaced gi:6522502.
                  IMPORTANT: This sequence is unfinished and does not necessarily
                  represent the correct sequence. Work on the sequence is in
                  progress and the release of this data is based on the understanding
                  that the sequence may change as work continues. The sequence may
                  be contaminated with foreign sequence from E.coli, yeast, vector,
                  phage etc. Order of segments is not known; 800 n's separate
                  segments. Contig_ID: 00034 Length: 5895bp
                  Contig_ID: 00089 Length: 12284bp
                  Contig_ID: 00176 Length: 15387bp
                  Contig_ID: 00207 Length: 6011bp
                  Contig_ID: 00545 Length: 14293bp
                  Contig_ID: 00757 Length: 3484bp
                  Contig_ID: 00760 Length: 6459bp
                  Contig_ID: 00804 Length: 3167bp
                  Contig_ID: 00883 Length: 3498bp
                  Contig_ID: 00960 Length: 5654bp
                  Contig_ID: 00967 Length: 7656bp
                  Contig_ID: 01108 Length: 9242bp
                  Contig_ID: 01342 Length: 11178bp
                  Contig_ID: 01408 Length: 10618bp
                  Contig_ID: 01440 Length: 2221bp.
                  * NOTE: This is a 'working draft' sequence.
                  * This record will be updated with the finished sequence
                  * as soon as it is available and the accession number will
                  * be preserved.
```

FEATURES

```
Location/Qualifiers
1..128247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25-26"
/clone="RP3-468K3"
/clone_lib="RPCI-3"
BASE COUNT 37843 a 22284 c 21565 g 35352 t 11203 others
ORIGIN
```

```
Query Match          9.6%; Score 19; DB 33; Length 128247;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 TTTTGAGATAATTAGACAA 140
    |||||
Db 120040 TTTTGAGATAATTAGACAA 120038
```

```

RESULT 18
AC015596 LOCUS
AC015596 Homo sapiens clone RP11-45A9, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC015596 HTG 17-NOV-1999
VERSION AC015596.1 GI:6447091
KEYWORDS HTG; HTGS_PHASE0; NULL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 135711)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collins, A., Doyle, M.,
Cooke, P., DeArnell, K., Dewar, K., Domino, M., Donelan, L., Gage, D.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galaan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieou, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Project Information
Center project name: L110
Center clone name: 45_A_9
-----
* NOTE: This record contains 154 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 854: contig of 854 bp in length
* gap of unknown length
*
* 855 1719: contig of 865 bp in length
* gap of unknown length
*
* 1720 2573: contig of 854 bp in length
* gap of unknown length
*
* 2574 3402: contig of 829 bp in length
* gap of unknown length
*
* 3403 4268: contig of 866 bp in length
* gap of unknown length
*
* 4269 5129: contig of 861 bp in length
* gap of unknown length
*
* 5130 5994: contig of 865 bp in length
* gap of unknown length
*
* 5995 6870: contig of 876 bp in length
* gap of unknown length
*
* 6871 7748: contig of 878 bp in length
* gap of unknown length
*
* 7749 8647: contig of 899 bp in length
* gap of unknown length
*
* 8648 9518: contig of 871 bp in length
* gap of unknown length
*
* 9519 10383: contig of 865 bp in length
* gap of unknown length
*
* 10384 11258: contig of 875 bp in length
* gap of unknown length
*
* 11259 12107: contig of 849 bp in length
* gap of unknown length
*
* 12108 12961: contig of 854 bp in length
* gap of unknown length
*
* 12962 13808: contig of 847 bp in length
* gap of unknown length
*
* 13809 14687: contig of 879 bp in length
* gap of unknown length
*
* 14688 15544: contig of 857 bp in length
* gap of unknown length
*
* 15545 16427: contig of 883 bp in length
* gap of unknown length
*
* 16428 17287: contig of 860 bp in length
* gap of unknown length
*
* 17288 18139: contig of 852 bp in length
* gap of unknown length
*
* 18140 18995: contig of 856 bp in length
* gap of unknown length
*
* 18996 19839: contig of 844 bp in length
* gap of unknown length
*
* 19840 20691: contig of 852 bp in length
* gap of unknown length
*
* 20692 21539: contig of 848 bp in length
* gap of unknown length
*
* 21540 22401: contig of 861 bp in length
* gap of unknown length
*
* 22401 23265: contig of 865 bp in length
* gap of unknown length
*
* 23266 24137: contig of 872 bp in length
* gap of unknown length
*
* 24138 25042: contig of 905 bp in length
* gap of unknown length
*
* 25043 25939: contig of 897 bp in length
* gap of unknown length
*
* 25940 26843: contig of 904 bp in length
* gap of unknown length
*
* 26844 27743: contig of 900 bp in length
* gap of unknown length
*
* 27744 28649: contig of 906 bp in length
* gap of unknown length
*
* 28650 29533: contig of 884 bp in length
* gap of unknown length
*
* 29534 30445: contig of 912 bp in length
* gap of unknown length
*
* 30446 31316: contig of 871 bp in length
* gap of unknown length
*
* 31317 32236: contig of 920 bp in length
* gap of unknown length
*
* 32237 33110: contig of 874 bp in length
* gap of unknown length
*
* 33111 33972: contig of 862 bp in length
* gap of unknown length
*
* 33973 34832: contig of 860 bp in length
* gap of unknown length
*
* 34833 35743: contig of 911 bp in length
* gap of unknown length
*
* 35744 36575: contig of 832 bp in length
* gap of unknown length
*
* 36576 37454: contig of 879 bp in length
* gap of unknown length
*
* 37455 38318: contig of 864 bp in length
* gap of unknown length
*
* 38319 39201: contig of 883 bp in length
* gap of unknown length
*
* 39202 40048: contig of 847 bp in length
* gap of unknown length
*
* 40049 40918: contig of 870 bp in length
* gap of unknown length

```



```
* * 40919 41800: contig of 882 bp in length
* * gap of unknown length
* * 41801 42686: contig of 886 bp in length
* * gap of unknown length
* * 42687 43548: contig of 862 bp in length
* * gap of unknown length
* * 43549 44392: contig of 844 bp in length
* * gap of unknown length
* * 44393 45271: contig of 879 bp in length
* * gap of unknown length
* * 45272 46139: contig of 868 bp in length
* * gap of unknown length
* * 46140 47065: contig of 926 bp in length
* * gap of unknown length
* * 47066 47930: contig of 865 bp in length
* * gap of unknown length
* * 47931 48798: contig of 868 bp in length
* * gap of unknown length
* * 48799 49684: contig of 886 bp in length
* * gap of unknown length
* * 49685 50537: contig of 853 bp in length
* * gap of unknown length
* * 50538 51400: contig of 863 bp in length
* * gap of unknown length
* * 51401 52280: contig of 880 bp in length
* * gap of unknown length
* * 52281 53155: contig of 875 bp in length
* * gap of unknown length
* * 53156 53999: contig of 844 bp in length
* * gap of unknown length
* * 54000 54846: contig of 847 bp in length
* * gap of unknown length
* * 54847 55691: contig of 845 bp in length
* * gap of unknown length
* * 55692 56570: contig of 879 bp in length
* * gap of unknown length
* * 56571 57443: contig of 873 bp in length
* * gap of unknown length
* * 57444 58326: contig of 883 bp in length
* * gap of unknown length
* * 58327 59202: contig of 876 bp in length
* * gap of unknown length
* * 59203 60078: contig of 876 bp in length
* * gap of unknown length
* * 60079 60939: contig of 861 bp in length
* * gap of unknown length
* * 60940 61811: contig of 872 bp in length
* * gap of unknown length
* * 61812 62684: contig of 873 bp in length
* * gap of unknown length
* * 62685 63588: contig of 904 bp in length
* * gap of unknown length
* * 63589 64454: contig of 866 bp in length
* * gap of unknown length
* * 64455 65322: contig of 868 bp in length
* * gap of unknown length
* * 65323 66187: contig of 865 bp in length
```

```
Query Match 9.6% Score 19; DB 43; Length 135711;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 93 AAAAAATATATATTTGGTG 111
|||||
Db 108911 AAAAAATATATATTTGGTG 108929
```

```
RESULT 19
HSJ709L21 DNA HTG 03-DEC-1999
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP4-709L21 map q42.13-43, ***
SEQUENCING IN PROGRESS ***, In unordered pieces.
```

```
ACCESSION AL078646
VERSION 20
KEYWORDS GI:6523750
SOURCE HTGS_PHASE1.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152583)
Donnelly, S.

Direct Submission
Submitted (03-DEC-1999), Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk
On Dec 3, 1999 this sequence version replaced gi:6491758.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known. 800 n's separate
segments. Contig_ID: 00157 Length: 1039bp
Contig_ID: 00362 Length: 1040bp
Contig_ID: 00988 Length: 1966bp
Contig_ID: 01010 Length: 1191bp
Contig_ID: 01548 Length: 124660bp
Contig_ID: 01722 Length: 1576bp
Contig_ID: 02108 Length: 4631bp
Contig_ID: 02181 Length: 1796bp
Contig_ID: 02245 Length: 1385bp
Contig_ID: 02591 Length: 2189bp
Contig_ID: 02919 Length: 1067bp
Contig_ID: 03086 Length: 1043bp.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source

Location/Qualifiers
1..152583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.13-43"
/clone="RP4-709L21"
/clone_lib="RPC1-4"

BASE COUNT 35861 a 31061 c 33040 g 43820 t 8801 others
ORIGIN

Query Match 9.6% Score 19; DB 33; Length 152583;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 TTTCGAGATAATTAGACAA 140
|||||

Db 6696 TTTCGAGATAATTAGACAA 6714

RESULT 20
AC009427

LOCUS AC009427 154129 bp DNA HTG 22-AUG-1999
DEFINITION Homo sapiens clone 45_M_3, *** SEQUENCING IN PROGRESS ***, 23
unordered pieces.

ACCESSION AC009427
VERSION AC009427.1 GI:5757602
KEYWORDS HTGS_PHASE1.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154129)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone 45_M_3
Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 154129)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baker,J., Balgwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
 Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
 Hagos,B., Headford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., McDonaid,P.,
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
 Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
 Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
 Direct Submission
 Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

TITLE
JOURNAL
COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 185: contig of 185 bp in length
 * gap of unknown length
 * 186 2563: contig of 2378 bp in length
 * gap of unknown length
 * 2564 5880: contig of 3317 bp in length
 * gap of unknown length
 * 5981 8528: contig of 2648 bp in length
 * gap of unknown length
 * 8529 10743: contig of 2215 bp in length
 * gap of unknown length
 * 10744 13869: contig of 3126 bp in length
 * gap of unknown length
 * 13870 16720: contig of 2851 bp in length
 * gap of unknown length
 * 16721 21145: contig of 4425 bp in length
 * gap of unknown length
 * 21146 25413: contig of 4268 bp in length
 * gap of unknown length
 * 25414 31523: contig of 6110 bp in length
 * gap of unknown length
 * 31524 37054: contig of 5531 bp in length
 * gap of unknown length
 * 37055 41745: contig of 4691 bp in length
 * gap of unknown length
 * 41746 47233: contig of 5488 bp in length
 * gap of unknown length
 * 47234 55611: contig of 8378 bp in length
 * gap of unknown length
 * 55612 63950: contig of 8339 bp in length
 * gap of unknown length
 * 63951 71243: contig of 7293 bp in length
 * gap of unknown length
 * 71244 79316: contig of 8073 bp in length
 * gap of unknown length
 * 79317 88629: contig of 9313 bp in length
 * gap of unknown length
 * 88630 99510: contig of 10881 bp in length
 * gap of unknown length
 * 99511 115336: contig of 16026 bp in length
 * gap of unknown length
 * 115337 128158: contig of 12622 bp in length
 * gap of unknown length
 * 128159 149566: contig of 21408 bp in length

* * 149567 154129: contig of 4563 bp in length.
 * Location/Qualifiers
 1. 154129
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="45 M.3"
 /clone_lib="RPCI-11 Human Male BAC"

BASE COUNT 40465 a 35517 c 34138 g 42930 t 1079 others
 ORIGIN

Query Match 9.6% Score 19: DB 41: Length 154129;
 Best Local Similarity 100.0%; Pred. No. 12:
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ATTTACAAGACACAGTTTG 170
 |||||

Db 110173 ATTTACAAGACAGTTTG 110191

RESULT 21

AC005165 155164 bp DNA PRI 20-JUN-1998
 LOCUS Homo sapiens BAC clone RG242H14 from 7p15-p21, complete sequence.
 DEFINITION AC005165
 ACCESSION AC005165
 VERSION AC005165.1 GI:3242748
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 155164)
 Becker,M., Tin-Wollam,A. and Harrison,M.
 The sequence of Homo sapiens BAC clone RG242H14
 Unpublished (1998)
 2 (bases 1 to 155164)
 Waterston,R.
 Direct Submission

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 Submitted (20-JUN-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 SUBMITTED BY: WUGSC
 Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
<http://genome.wustl.edu/gsc>
<mailto:sapiens@watson.wustl.edu>

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone; and the assembly was
 confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/grb/chr7> , send
<mailto:edgreen@nhgri.nih.gov> , or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone RG242H14 is from a release of the human BAC library

CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
 VECTOR: pBelOBAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RG241I02, 200 bp overlap.
 Actual start of this clone is at base position 88412 of RG241I02;
 actual end is at 155164 of RG242H14.

FEATURES

Source	Location/Qualifiers	
	1..155164	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7p15-p21"	
	/clone_lib="CITB-HS-A"	
	/clone="RG242H14"	
STS	195..455	
	/db_xref="GI:435320"	
repeat_region	267..383	/rpt_family="MIR"
repeat_region	449..578	/rpt_family="MIR"
repeat_region	1139..1194	/rpt_family="(TGA)n"
repeat_region	1754..1805	/rpt_family="L2"
repeat_region	1900..2827	/rpt_family="L2"
STS	complement(2701..3012)	
	/db_xref="GI:1916422"	
repeat_region	4372..4672	/rpt_family="Alu"
repeat_region	5769..5814	/rpt_family="L2"
repeat_region	6744..6897	/rpt_family="MIR"
repeat_region	8941..9194	/rpt_family="Alu"
repeat_region	9828..10013	/rpt_family="MER1_type"
repeat_region	10198..10511	/rpt_family="MER1_type"
repeat_region	11312..11486	/rpt_family="MIR"
repeat_region	11916..12524	/rpt_family="L1"
repeat_region	16979..17281	/rpt_family="Alu"
repeat_region	18519..18641	/rpt_family="MIR"
repeat_region	18811..18986	/rpt_family="L2"
repeat_region	21003..21264	/rpt_family="L1"
repeat_region	24656..24824	/rpt_family="MaLR"
repeat_region	24825..25120	/rpt_family="Alu"
repeat_region	25178..25371	/rpt_family="MaLR"
repeat_region	25405..25468	/rpt_family="MIR"
repeat_region	26231..26303	/rpt_family="MIR"
misc_feature	complement(26404..26808)	
misc_feature	/note="ubiquitin/CEP52 pseudogene"	
repeat_region	27675..28202	/rpt_family="Retroviral"
		28936..29075
		/rpt_family="MER1_type"
		29086..29384
		/rpt_family="L1"
		29399..30562
		/rpt_family="MER1_type"
		30610..30827
		/rpt_family="MIR"
		32223..32520
		/rpt_family="Alu"
		32817..32872
		/rpt_family="MER4_group"
		32943..33227
		/rpt_family="Alu"
		33550..33618
		/rpt_family="L2"
		36813..37651
		/rpt_family="Retroviral"
		37747..39045
		/rpt_family="L1"
		39041..40642
		/rpt_family="L1"
		40870..41028
		/rpt_family="L2"
		41054..41327
		/rpt_family="Alu"
		41402..41617
		/rpt_family="L2"
		42224..42607
		/rpt_family="MaLR"
		44185..44314
		/rpt_family="MIR"
		45622..46028
		/rpt_family="L1"
		46296..46607
		/rpt_family="Alu"
		46635..47175
		/rpt_family="L1"
		47198..47305
		/rpt_family="MaLR"
		47306..47927
		/rpt_family="L1"
		47960..48063
		/rpt_family="L1"
		48063..48649
		/rpt_family="L1"
		49199..49240
		/rpt_family="L2"
		49664..49739
		/rpt_family="MIR"
		50037..50111
		/rpt_family="MIR"
		50469..50569
		/rpt_family="L2"
		50763..50814
		/rpt_family="L2"
		50869..51031
		/rpt_family="MIR"
		51061..51399
		/rpt_family="L2"
		52484..52593
		/rpt_family="MIR"
		53714..54005
		/rpt_family="Alu"
		54029..54149
		/rpt_family="Alu"
		54160..54360
		/rpt_family="L2"
		55960..56143
		/rpt_family="MER1_type"
		56687..56988
		/rpt_family="Alu"
		57413..57511

```

repeat_region      /rpt_family="MIR"
58556..58574
/rpt_family="MIR"
repeat_region      60264..60469

```

```

Query Match          9.6%; Score 19; DB 11; Length 155164;
Best Local Similarity 100.0%; Pred. No.12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 39 GTGTTCCTCAAAATTACAT 57
      |||||
Db 829 GTGTTCCTCAAAATTACAT 847

```

```

RESULT 22
AC011107/c          158454 bp      DNA      HTG      29-OCT-1999
LOCUS              Homo sapiens clone 115_I_3, *** SEQUENCING IN PROGRESS ***, 14
DEFINITION          Homo sapiens clone 115_I_3, *** SEQUENCING IN PROGRESS ***, 14
ACCESSION            AC011107
VERSION              AC011107.2 GI:6139162
KEYWORDS             HTG: HTGS_PHASE1
SOURCE               human.
ORGANISM             Homo sapiens

```

```

REFERENCE
AUTHORS             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE               1 (bases 1 to 158454)
JOURNAL              Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
                    Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Bouckgaiter,B.,
                    Brown,A., Castle,A., Collangelo,M., Collins,S., Collamore,A.,
                    Cooke,P., DeAerellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
                    Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
                    Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
                    Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
                    Lehoszky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
                    McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
                    Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                    Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
                    Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                    Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
                    Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

```

```

REFERENCE
AUTHORS             Direct Submission
TITLE               Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
                    Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL              On Oct 29, 1999 this sequence version replaced gi:6006221.
                    All repeats were identified using RepeatMasker: Smit, A.F.A. &
                    Green, P. (1996-1997)
COMMENT              http://ftp.genome.washington.edu/RM/RepeatMasker.html.
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 14 contigs. the true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.

```

```

1 1907: contig of 1907 bp in length
gap of unknown length
1908 6037: contig of 4130 bp in length
gap of unknown length
6038 8861: contig of 2824 bp in length
gap of unknown length
8862 15375: contig of 6514 bp in length
gap of unknown length
15376 20175: contig of 4800 bp in length
gap of unknown length
20176 28001: contig of 7826 bp in length
gap of unknown length
28002 34064: contig of 6063 bp in length

```

```

* 34065 42130: contig of 8066 bp in length
* gap of unknown length
* 42131 52935: contig of 10805 bp in length
* gap of unknown length
* 52936 66599: contig of 13664 bp in length
* gap of unknown length
* 66600 81871: contig of 15272 bp in length
* gap of unknown length
* 81872 100212: contig of 18341 bp in length
* gap of unknown length
* 100213 122923: contig of 22711 bp in length
* gap of unknown length
* 122924 158454: contig of 35531 bp in length.

```

```

FEATURES
source
1..158454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="115_I_3"
/clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 51026 a 28914 c 28662 g 49852 t
ORIGIN

```

```

Query Match          9.6%; Score 19; DB 42; Length 158454;
Best Local Similarity 100.0%; Pred. No.12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 92 GAAATATATATATATTTGGT 110
      |||||
Db 157003 GAAATATATATATTTGGT 156985

```

```

RESULT 23
AC007027/c          166925 bp      DNA      HTG      16-APR-1999
LOCUS              Homo sapiens clone DJ0832014, *** SEQUENCING IN PROGRESS ***, 3
DEFINITION          Homo sapiens clone DJ0832014, *** SEQUENCING IN PROGRESS ***, 3
ACCESSION            AC007027
VERSION              AC007027.2 GI:4586096
KEYWORDS             HTG: HTGS_PHASE1
SOURCE               human.
ORGANISM             Homo sapiens

```

```

REFERENCE
AUTHORS             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                    Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE               1 (bases 1 to 166925)
JOURNAL              Waterston,R.H.
                    The sequence of Homo sapiens clone
                    Unpublished
REFERENCE            2 (bases 1 to 166925)
AUTHORS             Waterston,R.H.
TITLE               Direct Submission
JOURNAL              Submitted (06-MAR-1999) Genome Sequencing Center, Washington
                    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                    MO 63108, USA
COMMENT              On Apr 16, 1999 this sequence version replaced gi:4371318.
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 3 contigs. The true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.

```

```

1 10160: contig of 10160 bp in length
10161 10178: gap of unknown length
10179 83304: contig of 73126 bp in length
83305 83322: gap of unknown length
83323 168925: contig of 83603 bp in length.
Location/Qualifiers
1..166925
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="DJ0832014"
BASE COUNT 53634 a 29738 c 31389 g 52128 t 36 others
ORIGIN

Query Match          9.6%; Score 19; DB 33; Length 166925;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ACTAGAAAAATATATAT 106
      |||||
Db 69586 ACTAGAAAAATATATAT 69568

RESULT 24
AC009863/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-573D15 map 3, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC009863
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174050)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelell,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lechoczky,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (03-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 3, 1999 this sequence version replaced gi:5822695.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1835
Center clone name: 573_D_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 139143 bases at least Q40
Consensus quality: 158720 bases at least Q30
Consensus quality: 168324 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 174050; sum-of-ctnigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-ctnigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3770: contig of 3770 bp in length
* gap of unknown length
* 3771 6298: contig of 2528 bp in length
* gap of unknown length
* 6299 9074: contig of 2776 bp in length
* gap of unknown length
* 9075 12526: contig of 3452 bp in length
* gap of unknown length
* 12527 15845: contig of 3319 bp in length
* gap of unknown length
* 15846 21494: contig of 5649 bp in length
* gap of unknown length
* 21495 27000: contig of 5506 bp in length
* gap of unknown length
* 27001 34431: contig of 7431 bp in length
* gap of unknown length
* 34432 43961: contig of 9530 bp in length
* gap of unknown length
* 43962 53530: contig of 9569 bp in length
* gap of unknown length
* 53531 66775: contig of 13245 bp in length
* gap of unknown length
* 66776 80832: contig of 14057 bp in length
* gap of unknown length
* 80833 93016: contig of 12184 bp in length
* gap of unknown length
* 93017 106543: contig of 13527 bp in length
* gap of unknown length
* 106544 127466: contig of 20923 bp in length
* gap of unknown length
* 127467 174050: contig of 46584 bp in length.
* Location/Qualifiers
* 1..174050
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="3"
* /map="3"
* /clone="RP11-573D15"
* /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 48307 a 38133 c 37894 g 49708 t 8 others
ORIGIN

Query Match          9.6%; Score 19; DB 44; Length 174050;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TAGAAAAATATATATTTG 108
      |||||
Db 70487 TAGAAAAATATATATTTG 70469

RESULT 25
AC012149/c
LOCUS
DEFINITION Homo sapiens chromosome 3q27 clone RP11-119E3, *** SEQUENCING IN
PROGRESS ***, 29 unordered pieces.
AC012149
VERSION AC012149, 1 GI:6091636
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 192823)
REFERENCE

```

AUTHORS	Muzoy,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K., Bodolia,B., Bouck,J., Bowles,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M., Hollaway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Licharge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,K., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S., Nash,S., Neilson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabba,M., Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
	Direct Submission Unpublished 2 (bases 1 to 192823) Worley,K.C.
TITLE JOURNAL REFERENCE AUTHORS JOURNAL	Submitted (21-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
	COMMENT * NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
FEATURES source	1 864: contig of 864 bp in length * 865 1826: contig of 962 bp in length * 1827 3148: contig of 1322 bp in length * 3149 3991: contig of 843 bp in length * 3992 4821: contig of 830 bp in length * 4822 5883: contig of 1062 bp in length * 5884 6754: contig of 871 bp in length * 6755 7893: contig of 1139 bp in length * 7894 8736: contig of 843 bp in length * 8737 9534: contig of 798 bp in length * 9535 10543: contig of 1009 bp in length * 10544 11959: contig of 1416 bp in length * 11960 12783: contig of 824 bp in length * 12784 13565: contig of 782 bp in length * 13566 14382: contig of 817 bp in length * 14383 15924: contig of 1542 bp in length * 15925 17543: contig of 1619 bp in length * 17544 17976: contig of 433 bp in length * 17977 20336: contig of 2360 bp in length * 20337 26022: contig of 5686 bp in length * 26023 32770: contig of 6748 bp in length * 32771 36814: contig of 4044 bp in length * 36815 45401: contig of 8587 bp in length * 45402 56484: contig of 11083 bp in length * 56485 77905: contig of 21421 bp in length * 77906 100224: contig of 22319 bp in length * 100225 127870: contig of 27846 bp in length * 127871 137857: contig of 29987 bp in length * 137858 157858: contig of 34966 bp in length. Location/Qualifiers 1. .192823 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3q27"
BASE COUNT ORIGIN	53592 a 42230 c 42416 g 54504 t 81 others Unfinished: dJ1099D15 Contig_ID: 03046 acc=AL035456 Length: 1505
	Unfinished: dJ1099D15 Contig_ID: 03046 acc=AL035456 Length: 1505

Oy	90 TAGAAAAATATATATATTG 108
	Db 106574 TAGAAAAATATATATATTG 106556
RESULT 26 HS1099D15/c LOCUS	HS1099D15 231949 bp DNA HTG 23-NOV-1999 DEFINITION Homo sapiens chromosome 20 clone RP5-1099D15, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
	ACCESSION AL035456 VERSION AL035456.23 GI:6018445 KEYWORDS HTG; HTGS_PHASE1. SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 231949) Blakey,S. Direct Submission Submitted (08-OCT-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humquerry@sanger.ac.uk On Oct 11, 1999 this sequence version replaced gi:6002127. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: dJ1099D15 Contig_ID: 00002 acc=AL035456 Length: 1092 bp Unfinished: dJ1099D15 Contig_ID: 00040 acc=AL035456 Length: 1301 bp Unfinished: dJ1099D15 Contig_ID: 00147 acc=AL035456 Length: 4081 bp Unfinished: dJ1099D15 Contig_ID: 00229 acc=AL035456 Length: 1803 bp Unfinished: dJ1099D15 Contig_ID: 00351 acc=AL035456 Length: 2366 bp Unfinished: dJ1099D15 Contig_ID: 00365 acc=AL035456 Length: 1251 bp Unfinished: dJ1099D15 Contig_ID: 00461 acc=AL035456 Length: 1369 bp Unfinished: dJ1099D15 Contig_ID: 00522 acc=AL035456 Length: 1508 bp Unfinished: dJ1099D15 Contig_ID: 00588 acc=AL035456 Length: 1153 bp Unfinished: dJ1099D15 Contig_ID: 00755 acc=AL035456 Length: 2228 bp Unfinished: dJ1099D15 Contig_ID: 00777 acc=AL035456 Length: 125952 bp Unfinished: dJ1099D15 Contig_ID: 00939 acc=AL035456 Length: 1071 bp Unfinished: dJ1099D15 Contig_ID: 01003 acc=AL035456 Length: 2707 bp Unfinished: dJ1099D15 Contig_ID: 01006 acc=AL035456 Length: 1815 bp Unfinished: dJ1099D15 Contig_ID: 01070 acc=AL035456 Length: 1492 bp Unfinished: dJ1099D15 Contig_ID: 01141 acc=AL035456 Length: 1218 bp Unfinished: dJ1099D15 Contig_ID: 01155 acc=AL035456 Length: 1986 bp Unfinished: dJ1099D15 Contig_ID: 01161 acc=AL035456 Length: 1469 bp Unfinished: dJ1099D15 Contig_ID: 01238 acc=AL035456 Length: 1845 bp Unfinished: dJ1099D15 Contig_ID: 01487 acc=AL035456 Length: 1366 bp Unfinished: dJ1099D15 Contig_ID: 01496 acc=AL035456 Length: 1588 bp Unfinished: dJ1099D15 Contig_ID: 01886 acc=AL035456 Length: 1330 bp Unfinished: dJ1099D15 Contig_ID: 02082 acc=AL035456 Length: 2237 bp Unfinished: dJ1099D15 Contig_ID: 02171 acc=AL035456 Length: 1520 bp Unfinished: dJ1099D15 Contig_ID: 02387 acc=AL035456 Length: 1793 bp Unfinished: dJ1099D15 Contig_ID: 02398 acc=AL035456 Length: 2266 bp Unfinished: dJ1099D15 Contig_ID: 02481 acc=AL035456 Length: 1406 bp Unfinished: dJ1099D15 Contig_ID: 02511 acc=AL035456 Length: 12212 bp Unfinished: dJ1099D15 Contig_ID: 02801 acc=AL035456 Length: 2618 bp Unfinished: dJ1099D15 Contig_ID: 02962 acc=AL035456 Length: 4740 bp Unfinished: dJ1099D15 Contig_ID: 02963 acc=AL035456 Length: 1924 bp Unfinished: dJ1099D15 Contig_ID: 03015 acc=AL035456 Length: 3036 bp Unfinished: dJ1099D15 Contig_ID: 03046 acc=AL035456 Length: 1505

Query Match Best Local Similarity 100.0%: Pred.No. 11; Matches 19: Conservative 0; Mismatches 0; Indels 0; Gaps 0;	9.6%: Score 19: DB 43: Length 192823;

bp Unfinished: dj1099d15 Contig_ID: 03081 acc=AL035456 Length: 1537 bp Unfinished: dj1099d15 Contig_ID: 03084 acc=AL035456 Length: 1084 bp Unfinished: dj1099d15 Contig_ID: 03243 acc=AL035456 Length: 7634 bp Unfinished: dj1099d15 Contig_ID: 03255 acc=AL035456 Length: 2178 bp.
* NOTE: This is a "working draft" sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source

1. .231949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP5-1099D15"
/clone_lib="RPCI-5"
BASE COUNT 58430 a 42805 c 43044 g 58039 t 29631 others
ORIGIN
127 AGATAATTAGACAAGACAG 145
|||||
Db 35078 AGATAATTAGACAAGACAG 35060

Query Match 9.6%; Score 19; DB 32; Length 231949;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27

RPXX04/c
LOCUS 237523 bp DNA BCT 11-NOV-1998
DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment 4/4.
ACCESSION AJ235273
VERSION AJ235273.1 GI:3861237
KEYWORDS complete genome; prowazekii.
SOURCE Rickettsia prowazekii.
ORGANISM Rickettsia prowazekii.
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia.
REFERENCE 1 (bases 1 to 237523)
ANDERSSON, S.G., Zomorodipour, A., ANDERSSON, J.O.,
SICHERITZ-PONTEN, T., ALMARK, U.C., PODOWSKI, R.M., NASLUND, A.K.,
ERIKSSON, A.S., WINKLER, H.H. and KURLAND, C.G.
The genome sequence of Rickettsia prowazekii and the origin of
mitochondria
NATURE 396 (6707), 133-140 (1998)
99039499
2 (bases 1 to 237523)
ANDERSSON, S.G.E.
Direct Submission
Submitted (11-NOV-1998) S.G.E. ANDERSSON,
Siv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University
of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

FEATURES

source

1. .237523
/organism="Rickettsia prowazekii"
/strain="Madrid E"
/db_xref="taxon:782"
164. .1123
/gene="RP701"
164. .1123
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAA15137.1"
/db_xref="GI:3861238"
/translation="MAEPIFTLFGHKORILDRNNQKALNKLHIPIQVKNESHIS
DKERIFSELEKSYKECKIPFIQISNNVKKPIFRAODLINAKNLNITFDKYN
ALPQNSEKYSWKQIIEKRVDFVFTNEAQNLSIADGIVPKDKTTITDIRLYSVFN
NLINHYETHNGLSSTIYKVKDLIKTAKNODSKSIEITWPIISVDEATNLITAKGITS

gene

CDS

gene

CDS

EDQIYGLKLEIHEILKDPNNAENFKYVSQISROFKDKLDKQKHNDSNFNTOKVI
NDSNKDQVEHITSYVODKEQKQKPGQFLTRIFNFKIVFKQ"
2074. .3348
/gene="RP702"
2074. .3348
/gene="RP702"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAA15138.1"
/db_xref="GI:3861239"
/translation="MTTILVITIIIMISLSALFAATETAITASSPGKIHKLKAGNKR
AKTVLEVLKKEKVKVGLLIGNSLINTICTTATTTLISLGNNGTIVASGYMAFIIL
VFAEYVPKAIYAKVPEQLTILKMSSTIVIFLKLKPINIALDXTKEFCIFPINLPO
ISGTEVGRVIERHQEGGVKSKERNMGLGILDIRNMTVSEIMHRSNIIALNIDLPY
ELIITLISGAHTRIPLWONDRNIIIGLNLKDLKALYENNDKKVDKDFLLPWP
FIIDNALVQDLHAFERNHNFACVDEYGTGLGIIITLEDVIEEIVGPIITDEHRLNN
EIIQSNTEFIIGKTTTIRINDRELNDNLSDEDANTIAGLIIHKKIARIAPNOGEVAKIF
NFKIILKKIANKIDSVKITVLPMTETISSE"
3341. .5353
/gene="RP703"
3341. .5353
/gene="RP703"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF
(ccmf)"
/protein_id="CAA15139.1"
/db_xref="GI:3861240"
/translation="MSKIGNFLTLTMCLSGILTLIIIPWLDYLIKKHNNICLKFF
PWICTSOVMTQAITFASCLCSIFACETLVYAFITSDFSVQNVFHSHTLPLIKI
SGSSHEGMLMFWCLLIQVSCYIFCLDEKMLKFSIIILSAIQLLFSSIFLSTN
PEHMSFYPKEGLNPLQDIALSIHPPLLYLGYSVAPPTIVCAIFLAIARLD
NNLPGYFTGYKIQTSIDRISIRLLKIFACGILFTTIGISLGMWAYRELNMGILWF
FDPEVENSILFPLWSLIMLHSHIIVTKINRMOHWITLISITFLVYFSTFLVRSQFI
NISHAFSSERWIVLVLICIGIGLYVIRLNCSEKTIYNTFNNTNCSNMT
IKEKAILIGNIFELSLILICATLYPTIVALENDKPIIINEEFFKNFLIFVIPILC
TIGLFTNSSIKKHITIMLSLITVILLIKFSLVSLITMTSIFLHNVHLLV
KINFRNSLRKASSIMLGHFGVALIAFSTIMNSLQSEMDFICEVGTSEKTNARAYT
LONIKFAQKNYRQIAEFNLEDHKNNTIILKPNFYIIEQKLSQESDIHSLIYDL
YAVLSNIDGKIIHAKIYKPMMSFIWIGVILTITAGFVIALIRKNSY"
/gene="RP704"
/complement(7264. .12195)
complement(7264. .12195)
/gene="RP704"
/complement(7264. .12195)
/codon_start=1
/transl_table=11
/product="CELL SURFACE ANTIGEN (sca5)"
/protein_id="CAA15140.1"
/db_xref="GI:3861241"
/translation="MAQPNFLKKIISAGLVTAATATIVAGFSVGAAMQYNRRTTN
AAATFDGIGFDOAAGANI PVAPNSVITANANNPITFNPNGHLSLFDLTANDLAVT
INEDTLGFIINIAQAKFFNTVAACKIILNITGOGITVOEASNTINAQALTKVHG
AIIANDLSIGLSITFRAAPSLEFNLIPTQEARPLTGANSKIYNGNGTLINTG
FIQSDNFTAGIKIINDCCOGLMFNTPORANILNLQVGGNTINFGIDGKGLVLY
SKNGAATFNTGTGLGNLKGIIELNLAAGKLIISOGGAANAVIGTDNGAGRAAGFI
VSDNGNAATISGOVYAKNMVIOAGAGQVTFEHIIVGLGTTTFFKTSADSKVITE
NSNFGTNGNLDQIVVPDTKILKGNFIDGVKNNGTAGVITFNAGALVSASTDPN
IATVINIAEAGVAVELSGIHIAELRLNGSGIFKLADGTIVINGVONALMNNNA
LAAGSIQDLSAGIITGIDGNGVAAQLHITLANDASKILALDGANIIGANVGAATHE
QANGCTIKLTNNIVNFDLITTDKTVGDVASSLTNNQTLITINGSIGTVVANTKT
LAQALIGSKTILANGDVAINELVINGNSVOLNHTYLLTKTINANGOGLIIVAADP
LNIITLADGNLNSAENPLSTIHFAKKAANADSLNVGKYNLIYANNITTDANVGS
LHFRSGGTSVSGTVGGQGHKLNLLDNGTIVFLUGDTTFNGGTFKIGKSLQLSN
NTTDTHEVSADNTGTLEFNTDPIITVLNKGAGYFGLKQVILSGPGNIVFNEIGNV
IVHGAANSISFENASLSTSLPSPGLDVLTKSTVGNGTDFNFAIVVSGIDS
MINNGOIIIGDKRIIALSLGSDNISIVNANTLXSGITTKNOGTVTLSGMPNPGT
IYGLGLENGSKLKQVFTTDDYNNLGSIIANNVTINDYVLTITGGTAGTDFDAKITLG
SYNGNANRVDSVTFESOPSRMVAQKNTVYLGALVNSLSDTDPVASVRFTGN
DSGAGLQGNIVSQNIDFGTYNLTILNSNVLGGTTAINGEIDILLNLIIFANGTSTW
GNTSISLTLNSGNGVQVVADEAQNATTTGTTTIKIQDNANANFSGTQATLIQ
GARGNLTGAPNFAVTSNIFVKEIYETIIRDSNDYVLTIRTNVNLNANFSGTQATLIQ
APGVSONISRCLESTNTAAYNNMLLAKDPDVAITFVGAIAATD7SAAVTVTNNDTOKT

gene

CDS

```

OQLLSNRLGLRLYLNAETSDVAGSATGAVSSGDEAEVSYGVNAKPFYNIAREQKKG
IAGYKATYGVVGLDTLASDNLMIGAAIGITKTDIKKODYKKGDKTDINGLSFLYGV
SOQLVKNFEAQAAGNATLANKVSKSORVFEESNGKMSKIOIAAGNYDNMTFGNLI FG
DYNAMPNVLVTPMAGLSYLKSSNENYKGTGTIVANKRINSKPSDEVDLIVGAKVAGST
VNIIDIVIYEIHSIFSVVHVKNGLSNQSMQLGQAPFISQPDRTAKTSYINIGLSANI
KSDAKMEYGIGYDFNSASKYTAHQGTGLKVRVNF"
complement(12953..13216)
/gene="RP705"
/Note="Pseudogene that lacks ATG or GTG initiation codon.
Homology to guanosine-3,5-bis(diphosphate)
3-pyrophosphohydrolase."
complement(12953..13216)
/gene="RP705"
complement(13563..14507)
/gene="RP706"
complement(13563..14507)
/gene="RP706"
/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:3861241"
/translations="MKIRQAKPVIIGISGPDLTAERLPEEHPNPLGIILFRNRIRK
NEGEQKALIKLIADIKVEIGDNTIIAIDQEGWVKRLIAPTIFYDAPAAQTFTLQ
TKNYSTIAKREIGINDFAPVADLIHDGADKIIISDRSGKEPEIVPLFLSAIA
GLQEKYTIACIKHIPGHRATVDSHIELPDIIDTSKLTEDIDFKVKELAKYDIKLA
MTAHHVITALDPONPATLSQIVIDYIKSNIGFKGLIISDTIEMQALSGSMADITKGA
DAGVDIVLECTGFNMSEVLGSVAENSIDKFSDLFIS"
complement(14552..15025)
/gene="RP707"
complement(14552..15025)
/gene="RP707"
/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:3861243"
/translations="MLNNKLTKKYVASEVIRILNALHKLRYLEKIPGLSHYKINN
RKYTAHDIDLLQKFLFVSRREYLEVGYHSDSDYDLANSFNHCASDTRQNEFOG
KNSCIUNLHVQRQDSQSAVLUSDMDNTVSSRIDLLTNFNHNSIQIKILLIS"
complement(15220..15507)
/gene="RP708"
complement(15220..15507)
/gene="RP708"
/codon_start=1
/transl_table=11
/product="INTEGRATION HOST FACTOR ALPHA-SUBUNIT (hima)"
/db_xref="GI:3861243.1"
/translations="MTITKNKISLMSLKKLGFNSNNLCEAIVNTVFSNILEIAKVOKLT
LKNFGSEVRQKTPRGINFHTKSPVMIASKNLRFTPTSEKLKALINKSML"
18405..18701
/gene="RP709"
18405..18701
/gene="RP709"
/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:3861244"
/translations="MAVGROYIEDHTPEQLHKLCDNLVKYIELETKAITHKAAEDTEV
VIEVSLDLSDDASAIINKKATITVSDDIANSNEDAAARNTNKGSGDVLRCYQ"
complement(19685..20789)
/gene="RP710"
complement(19685..20789)
/gene="RP710"
/Note="Pseudogene with two authentic framehifts and four
authentic in-frame termination codons. Homology to
transposase."
complement(21112..21608)
/gene="RP711"
complement(21112..21608)
/gene="RP711"

```

```

/Note="Pseudogene with one authentic frameshift. Homology
to resolvase."
complement(22602..23066)
/gene="RP712"
complement(22602..23066)
/gene="RP712"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAA15145.1"
/db_xref="GI:3861245"
/translations="MSIEKEFWASHKQVKEIGGTNLNGKHLDFKAVAYTSDDIEA"

Query Match      9.6%; Score 19; DB 1; Length 237523;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 TGAAGAATCTATTATATG 194
|||||
Db 153913 TGAAGAATCTATTATATG 153895

RESULT 28
BNASLRGA
LOCUS BNASLRGA 1563 bp mRNA PLN 25-JUL-1995
DEFINITION Brassica napus S-locus related protein (SLR-WS-1) mRNA, complete
cds.
ACCESSION L10737
VERSION L10737.1 GI:484111
KEYWORDS
SOURCE Brassica napus (strain Westar, sub_species oleifera) cDNA to mRNA.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Brassica.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Robert L.S., Allard, S., Franklin, T.M. and Trick, M.
TITLE Sequence and expression of endogenous S-locus glycoprotein genes in
self-compatible Brassica napus
JOURNAL Mol. Gen. Genet. 242 (2), 209-216 (1994)
MEDLINE 94211211
REFERENCE 2 (bases 1 to 1563)
AUTHORS Robert L.S.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1994) Laurian S. Robert, Agriculture Canada,
Plant Research Centre, Ottawa, Ontario, K1A 0C6, Canada
FEATURES
source
1..1563
/organism="Brassica napus"
/strain="Westar"
/sub_species="oleifera"
/db_xref="taxon:3708"
mRNA 1..1563
/gene="SLR-WS-1"
gene 1..1563
/gene="SLR-WS-1"
CDS 14..1354
/gene="SLR-WS-1"
/codon_start=1
/product="S-locus related protein"
/protein_id="AAA70399.1"
/db_xref="GI:484112"
/translations="MKGVIPHYHSYFFVILVLPFAFSTNTLSSNCALTISSNKT
LYSPGDFELGFFKTKNSQDSDTRWILGWIKTISDKRTYVWVANDNPLHNSIG
TLKSHASVLDDSDTVMSTNLGTVAHLPFAELLANGFVLRDSKNTDLDRFMQ
SFDYPTDPLLEPMKIGRKSSEKELITLTSWKSPDPSGDSLSILETGFLEHLYFL
RMSKTYRTGPNVGRFNGPKKQMSYIDNSFIDNNEVAYTFKVNHNHNSHRE
KMSGTGLQVITWTITVPQRNMFWSFPEDCDLYKVGYPAYCDMDHTSPTCNKIGFV
PKNAGRWDLRDMGGCVRRSLKSGEDGFLRMSQNKLPETSEALVKRIGLKECREK
CYRDCNCTGYANMDMDGGSGCVTWGTGFLDMRKYDAGGODLYLVKVAESLYPS"
BASE COUNT 483 a 306 c 349 g 425 t
ORIGIN

```


Query Match 9.1%; Score 18; DB 7; Length 1563;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 GAAAAATATATATTGG 109
|||||
Db 1522 GAAAAATATATATTGG 1539

RESULT 29
BNSLOGPB
LOCUS BNSLOGPB 1563 bp mRNA PLN 29-JUL-1997
DEFINITION B.napus mRNA for endogenous S-locus glycoprotein (cone SLR1-Ws-1).
ACCESSION 221609
VERSION 221609.1 GI:17817
KEYWORDS endogenous S-locus glycoprotein gene.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 1563)
AUTHORS Oldknow,J., Franklin,T.M., Allard,S., Robert,L.S. and Trick,M.
TITLE DNA sequences of the two homologous SLR1 genes of Brassica napus
JOURNAL cv inestar
SEX Plant Reprod. 8, 254-255 (1995)
REFERENCE 2 (bases 1 to 1563)
AUTHORS Robert,L.S.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1993) Robert L. S., Agriculture Canada, Plant
Research Centre, 960 Carling Avenue, Ottawa, Ontario, Canada, K1A
0C6

FEATURES
source Location/Qualifiers
1..1563
/organism="Brassica napus"
/strain="Westar"
/db_xref="taxon:3708"
/clone_lib="lambda gt 10"
/clone="SLR1-Ws-1"
14..1354
/codon_start=1
/product="glycoprotein"
/protein_id="CAA79735.1"
/db_xref="GI:2285899"
/translation="MKGVIPVHHSTYFFVILVLPFAFSTNTLSSNEALTSSNKT
LVSPGVFELGFEKATKNSQDSTDRWYLGWYKTTSDKRTYVWVNRDNLPHNSIG
TLKISHASVLDDQSTTVWSTNLGVAHLPVTAELLANGFVLRDSTNLDLRFMQ
SDYPVDVTLLEPMKIGRNLIGSENEKILTSKSPDPSGDYSLILETEGFLHEFYLF
KNEFKYRTGPNVGRFNGIPKMQNWSYIDNSFDNNEEVAYTFKVHNNNMTHSRF
RMSTGYLQVITWTKTPQRMFWSPEDTCDLYKVCGPYACDMHTSPTCNCIKGEV
PKNAGRDLRDMSGGCVRSKSCGEGDFLRMSQKLPTSEALVERKIRIGLKECREK
CVRDCNCTGYANNDIMDGGGCVVTIGELVDMRKYDAGGQDLYVKVAEASLVPS"

CDS
BASE COUNT 483 a 306 c 349 g 425 t
ORIGIN

Query Match 9.1%; Score 18; DB 7; Length 1563;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 GAAAAATATATATTGG 109
|||||
Db 1522 GAAAAATATATATTGG 1539

RESULT 30
BNSLIR
LOCUS BNSLIR 1794 bp DNA PLN 09-FEB-1996
DEFINITION B.napus (Westar) SLR1 gene for SLR1 glycoprotein.
ACCESSION 226914

VERSION 226914.1 GI:456316
KEYWORDS SLR1 gene; SLR1 glycoprotein.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 1794)
AUTHORS Oldknow,J.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1993) Oldknow J., Cambridge Laboratory, Institute
of Plant Science Research, Brassica & Oilseeds Research, John Innes
Centre, Colney Lane, Norwich, Norfolk, UK, NR4 7UJ
REFERENCE 2 (bases 1 to 1794)
AUTHORS Oldknow,J.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1994) Oldknow J., Cambridge Laboratory, Institute
of Plant Science Research, Brassica & Oilseeds Research, John Innes
Centre, Colney Lane, Norwich, Norfolk, UK, NR4 7UJ
REFERENCE 3 (bases 1 to 1794)
AUTHORS Oldknow,J., Franklin,T.M., Allard,S., Robert,L.S. and Trick,M.
TITLE DNA sequences of the two homologous SLR1 genes of Brassica napus
JOURNAL cv inestar
SEX Plant Reprod. 8, 254-255 (1995)
COMMENT On Feb 23, 1994 this sequence version replaced gi:407774.
FEATURES
source Location/Qualifiers
1..1794
/organism="Brassica napus"
/strain="Westar"
/db_xref="taxon:3708"
/clone="2a2"
/tissue_type="Stigma"
/clone_lib="Lambda gt10"
142..1476
/gene="SLR1"
142..1476
/gene="SLR1"
/citation=[3]
/codon_start=1
/evidence=experimental
/product="SLR1 glycoprotein"
/protein_id="CAA81540.1"
/db_xref="GI:456317"
/translation="MRGVIPVHHSTYFFVILVLPFPHVSTNTLSPNENLTSSNKT
TLVSPGVFELGFEKTTTNSPDGDRWYLGWYKTTSGHRTYVWVNRDNLHNSMG
TLKISHASVLDDHSTNPWSTNFTGVAHLPVTAELLANGFVLRDSTNLDLRFMQ
SDYPVDVTLLEPMKIGRNLIGSENEKILTSKSPDPSGDYSLILETEGFLHEFYLF
KNEFKYRTGPNVGRFNGIPKMQNWSYIDNSFDNNEEVAYSFQVNNNNHHIRFRM
SSTGYLQVITWTKTPQRMFWSPEDTCDLYKVCGPYACDMHTSPTCNCIKGEV
NAGRDLRDMSGGCVRSKSCGEGDFLRMSQKLPTSEAVVDKRIGLKECREK
RDCNCTGYANNDIMMNGGCGVMTGELDMRKYNAGGQDLYVKVAAASLVPS"
142..225
sig_peptide
mat_peptide
polyA_signal
BASE COUNT 566 a 351 c 381 g 496 t
ORIGIN

Query Match 9.1%; Score 18; DB 7; Length 1794;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 GAAAAATATATATTGG 109
|||||
Db 1642 GAAAAATATATATTGG 1659

RESULT 31

```

MMU13839      1825 bp      mRNA      ROD      03-DEC-1998
LOCUS          Mus musculus vacuolar adenosine triphosphatase subunit C mRNA,
DEFINITION    complete cds.
ACCESSION     U13839
VERSION       U13839.1 GI:3955097
KEYWORDS      house mouse.
SOURCE        Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 1825)
AUTHORS      Howell, M.L. and Dean, G.E.
JOURNAL      Unpublished
TITLE        cDNA sequences for mouse vacuolar ATPase subunits
REFERENCE     2 (bases 1 to 1825)
AUTHORS      Dean, G.E.
JOURNAL      Direct Submission
TITLE        Submitted (19-AUG-1994) Gary E. Dean, Mol. Genetics, Biochemistry,
              and Microbiology, University of Cincinnati College of Medicine, 231
              Bethesda Avenue, Cincinnati, OH 45267-0524, USA
FEATURES     Location/Qualifiers
              1..1825
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /cell_line="ES-D3"
                /dev_stage="18.5 day"
                156..1304
                  /function="vacuolar acidification"
                  /note="subunit C, V-ATPase"
                  /codon_start=1
                  /evidence=experimental
                  /product="vacuolar adenosine triphosphatase subunit C"
                  /protein_id="AAC83084.1"
                  /db_xref="GI:3955098"
                  /translation="MTEFLISAPGKTCQOTWEKLAATKNNLAVSKENIPDLK
                    VGLDLVGLSDELAIDAEVGVKQVAGYADLEDSKQVQENLLASGYDLITYI
                    TRQDWMAKPIPKSLKNISEIIAKGVTQIDNDLKSASAYNNLKNLQNLKKNAGS
                    LLTSLAEIKKDFDVLDSYLVTLVLPVKLNNDWIKQYETLAEMVPRSSNVLSE
                    DQSYLCNVTLFKKAVDDFRHAKENKFIVRFQYNEEKMAKDEMTLELSDKKQF
                    GPLVRLKVFSEAFIAWIIKALRVFESVLRGLPVNFQAMLLQPNKSVKKLREV
                    LHELYKHLDSAAAIIDAPMDIPGLNLSQGEYYPVYVYKIDCNLEFFK"
BASE COUNT   521 a 371 c 434 g 499 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

REFERENCE     2 (bases 1 to 2139)
AUTHORS      Trick, M.
TITLE        Genomic sequence of a Brassica S locus-related gene
JOURNAL      Plant Mol. Biol. 15 (1), 203-205 (1990)
MEDLINE      91355861
COMMENT      Data kindly reviewed (02-JUN-1990) by Trick M.
FEATURES     Location/Qualifiers
              1..2139
                /organism="Brassica oleracea"
                /strain="alaboglabra"
                /db_xref="taxon:3712"
                /cell_line="plant line S63"
                /clone.lib="Lambda EMBL4"
                /clone="BS63-1"
                220..226
                  /note="TATA box"
                329..335
                  /note="TATA box related"
                379..413
                  /note="S locus related glycoprotein precursor (AA -28 to 416)"
                /codon_start=1
                /protein_id="CAA36307.1"
                /db_xref="GI:17897"
                /translation="MRGVPINHYHSYTLFFVLVLPFHVFTNTLSPNEALTISNNK
                  TLVSPGDVFLGFEKFTTRNSPDGTRWYLGIVYKTTSGHRTYVVAANRDNALHNSMG
                  TLKISHASLVLLDHSNTFVNSTFTVAHLPTVAELLANGFVLURDSKTLNDLDRFMQ
                  SFDPVDTLLPEMKLGRNLGSENEKILTSWKSPDPSGDFSLTEGLHFEFYLL
                  KNEKVVYTPGNGVRENGIPKMONWSYIDNSFIDNNEVAYSFOVANNHNIHTRFM
                  SSTGLQVITWTKIVPORNFWFSPEDTCDLYKVGPIAYCDHHSITCNCIKGVFVK
                  NAGRWDLRDMSGGSCSKSGEGGDFLRMSOMKLPETSEAVVDRKRGILKREKCV
                  RDCNCTGYANNDIMNGSGCVMTGELDDMRKYNAGGODLYVKVAAASLVPS"
                379..462
                  /product="signal peptide (AA -28 to -1)"
                463..1710
                  /product="mature peptide (AA 1 to 416)"
                1775..1780
                  /note="polya signal"
BASE COUNT   680 a 409 c 446 g 604 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```



```

exon      2288..2678
/feature="rpl15"
/citation={2}
/number=4
/evidence=experimental
3'UTR     2519..2678
/feature="rpl15"
/citation={2}
/evidence=experimental
polyA_site 2678..2680
/citation={2}
/evidence=experimental
BASE COUNT 920 a 522 c 550 g 992 t
ORIGIN

Query Match      9.1%; Score 18; DB 7; Length 2984;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 AAAATATATATATGGTG 111
|||||
Db 798 AAAATATATATATGGTG 781

RESULT 35
LOCUS      AF102623 7186 bp DNA BCT 24-MAY-1999
DEFINITION Methanosarcina barkeri dimethylamine corrinoid protein Mtbc (mtbc),
            trimethylamine methyltransferase Mtcb (mttb), trimethylamine
            corrinoid protein Mtcc (mttc), putative transmembrane protein Mttp
            (mttp), and dimethylamine methyltransferase Mtbb1 (mtbb1) genes,
            complete cds.
ACCESSION  AF102623
VERSION     AF102623.1 GI:4262423
KEYWORDS   Methanosarcina barkeri.
SOURCE     Methanosarcina barkeri.
ORGANISM   Methanosarcina barkeri
            Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
            Methanosarcina.
REFERENCE  1 (bases 1 to 7186)
AUTHORS   Ferguson,D.J., Jr. and Krzycki,J.A.
TITLE     Reconstitution of trimethylamine-dependent coenzyme M methylation
            with the trimethylamine corrinoid protein and the isozymes of
            methyltransferase II from Methanosarcina barkeri
JOURNAL   J. Bacteriol. 179 (3), 846-852 (1997)
MEDLINE   97158682
REFERENCE  2 (bases 1 to 7186)
AUTHORS   Ferguson,D.J., Gorlatova,N., Paul,L., Grahame,D. and Krzycki,J.A.
TITLE     The corrinoid protein from Methanosarcina barkeri specific for
            dimethylamine: CoM methyl transfer
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 7186)
AUTHORS   Paul,L. and Krzycki,J.A.
TITLE     The genes encoding the trimethylamine and dimethylamine
            methyltransferases of Methanosarcina barkeri are cotranscribed and
            interrupted by translationally bypassed in-frame amber codons
JOURNAL   Unpublished
REFERENCE  4 (bases 1 to 7186)
AUTHORS   Paul,L. and Krzycki,J.A.
TITLE     Direct Submission
JOURNAL   Submitted (27-Oct-1998) Microbiology, Ohio State University, 484
            West 12th Ave, Columbus, OH 43210, USA
FEATURES   Location/Qualifiers
            1..7186
            /organism="Methanosarcina barkeri"
            /strain="MS"
            /db_xref="taxon:2208"
            734..1375
            /gene="mtbc"
            734..1375
            /gene="mtbc"
            /function="participates in dimethylamine:coenzyme M methyl

```

```

transfer"
/feature="ORF identified by N-terminal sequence of isolated
            corrinoid protein"
/codon_start=1
/transl_table=11
/product="dimethylamine corrinoid protein Mtbc"
/protein_id="AAD14629.1"
/db_xref="GI:4262424"
/translation="MSKEELQLQELADAIISCKKDTVLAVVEKAKGELEPSEIEIKGLA
            AGMNEVGVRFERGKFLPHVMMADAMTAGAALKDLMPEGASGSLGVINGTVEGD
            VHDIKAIYSTMLOSAGFEVHDIGROVPIRNFIEKAKEVNADMIGISALMTTLQGOQ
            SVIELKEGLRDKVVMVGAPATQAWADKIGADCYAENATEAVAKARELLA"
1402..2889
/feature="mttb"
1402..2889
/feature="mttb"
/feature="UAG codon is translated by an undetermined
            mechanism; ORF identified with N-terminal sequence of 52
            kDa trimethylamine methyltransferase polypeptide"
/codon_start=1
/transl_except=(pos:2401..2403,aa:OTHER)
/transl_table=11
/evidence=experimental
/product="trimethylamine methyltransferase Mtcb"
/protein_id="AAD14630.1"
/db_xref="GI:4262425"
/translation="MAKNAVAGFNALGVNLFTTDELKATHYATMEVLMDDPGIOV
            SDPARQIFKENGCEVNEKTVVVKIPEYLRKALQALAPSRFVLWGRDKRNTVQECGG
            KYHCTGFGVKYQDQYVVDSEKDIADIAKCDWAENIDIFSLPVSARDING
            QGAQVHETLPLANTAKHFHIDPVGENVEYRDIVKAYIGGDEEERKRPFSMLL
            CPTPLSLVSNACQVLIIGARFGIPVNVLSMAMSGGSPVYLAGTLVTHNAEVLGIV
            LAQTVEGAKWYGSSTTFDLKKGTAPVSGPELGLISAVALKADYGLGSPYVAGSX
            SDKVPDQAGHEKMTLLPALAGANTYAGAGLELGMTFSEQLVIDIONDIFSVMVKX
            AMOGIPYSEETLAVESIQVIGNNPLAKTQVLDYPSNMLLDRHFGDWAAGS
            KDLATVAHEKVEDVLKKNQVTPIDADIFKDMAIVDKADKPRGM"
2920..3573
/feature="mttc"
2920..3573
/feature="mttc"
/function="participates in trimethylamine:coenzyme M
            methyl transfer with TMA methyltransferase"
/feature="ORF identified by N-terminal sequence of corrinoid
            protein"
/codon_start=1
/transl_table=11
/product="trimethylamine corrinoid protein Mtcc"
/protein_id="AAD14631.1"
/db_xref="GI:4262426"
/translation="MANKEETIAKAKEAITDFDDELAEEVANEALAAGIDPVLEIEKG
            FTAHEVEGKFGOGELFLPHVLAEEAMNSGKIVITPEMERKRSOTKSLGTVAIGTI
            EGDHISCKDIVASMLNATAGFVVDLGRDVPINTFEVKYKELKPOVVASSALMTVMV
            NQIQIEQLKEAGVDRDQVKTWVGAPVQDMADKIGADIYGESANDAVAKVRAALNV"
3885..4946
/feature="mttb"
3885..4946
/feature="mttb"
/function="possible methylamine permease"
/codon_start=1
/transl_table=11
/product="putative transmembrane protein Mttb"
/protein_id="AAD14632.1"
/db_xref="GI:4262427"
/translation="MEGVFVDLKLAKKQESKRRKISKGYMAFFCAIFGWIYLPGTVV
            WLNPFDEMSAIKTSKDGISLVVTAVLITAFNALAVMLAMVWNGVIGKYGELVRT
            LKEFPCSKWFFLASIFGGMAILFGFAMGFGGSFAAVAGLLYPVGSILAYVYWG
            EKISKRAIGCIATIVLOGISYGGGLFTELSSCNVPVIGYLAGLMAAGWTEGATAG
            KGLDIAPDAGLTIRFLGENTIIWIIIVPVLYVGPVMSFALOAPDPTIILLPAG
            IIFGCIYVTKRSFPLGVGVGOGIGNLIGLCAIIFLFLFGDVPWTIIIGALUCIA
            GSFIMFTEDASELETLRGE"
5470..6873
/feature="mtbb1"
5470..6873
/feature="mtbb1"

```

/note="UAG codon is translated by an undetermined mechanism: ORF identified by N-terminal sequence of 50 kDa DNA methyltransferase polypeptide"

/codon_start=1

/transl_except="(pos:6535..6537,aa:OTHER)

/transl_table=11

/evidence=experimental

/product="dimethylamine methyltransferase Mtbb1"

/protein_id="RAD14633.1"

/db_xref="GI:4262428"

/translation="MATEYALRMGDKRKYITKEKIVSEIEAGTADAADIGEIPALSA
NEMDLAELIMPCKTVSEQMEIPVTHDITGLDQDQSGVGPSSRLVGCMT
ERAFADMEICHIDYSPKPKVSVNEQAMEYCOQNMVPIPLFGAMPNMGILYTPD
GFPEPGLMKAFAKIQEAWESMEHAHELTRDTVMVKLFLFASGADGVNFTGAAGD
GMYCTLHAICALRKEFPDMYIEAGMAGECVLGMHGNLDYDGVTLAGLWPHQOALVA
KAMVFCFPCVNTSTKTSAMNLARAVTFMKAAYEASPICHVDVGMGVGJEMLETP
PIDAVTRASKARVEIAGVDGIXIGVDPDLPMPHIAHNASGNTGNRAAGDLVAREFSK
NNRIGEAKEYVAKLIGVDKMDLVDHVMRELREEDLDIGIITSVPGAAGIAAKWNIEK
LLDIKINSCNFRKOIA"

BASE COUNT 2127 a 1507 c 1692 g 1860 t

ORIGIN

Query Match 9.1%; Score 18; DB 2; Length 7186;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TAGAAAAATATATATT 107

|||||

Db 394 TAGAAAAATATATATT 411

RESULT 36

MXH99740/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

/clone="RI-5.6"
/clone="HIII-13"
/clone="opp2.1"
631..3516
/gene="P100"
631..3516
/gene="P100"
/codon_start=1
/transl_table=4
/product="intrinsic membrane protein"

/protein_id="CAA68073.1"
/db_xref="GI:1487982"
/db_xref="SPTREMBL:Q49551"
/translation="MTKSKLWLLGLGLSLAVGVLVAASCKRIDPAYVKGEVYEFN
AVNPEFIDADVSSYGYINSLPATYVSGELVRKQALTOTKTIITEKETKTTTTL
VKKPKAYELAEALVITEKNTKOVFSDSELPSEYQPTKTVTDINGKEYVY
EKTNVYTKGDKNVEKRSINSAFFLDKLANAKLQFVIRKGVWVHDGDKGIEYK
AKDFWVLRITYSRNLEYRUNTGGSKDELAKSKLAENSPLYGEOMSGNIYVAL
FGIDANKLFKEDKFIQVSAEGLNKDREAITFESINGEKDAMISFEFSQFASGYD
FMPAPSDYIDALNAKNEQVLKSNGLALSKESEAYNALNSMDKSLQYKIQGYWG
TSKSTLYSGYYNKGTEIQKLPNTLYWDRKVEDDSTIKTIVNFOSKAVEPAL
YNOTAFKYSOGTISOIGFSSLTEPOKKEVSSEPGKYGLKNERLNVQTSIFGVRT
FISKLGKYNENENFAKLYGSSLAIDSGKADTSESYLEGKGLSRTILLYAAVND
TYAQAASSGQATHMLAKVPEGSNLGGSDGQGTAKLNTITDFDQANSFLDKDQNKVD
FGDKIKELSPENAKASETSEISERKMSAGFKVLEKMKLELDEFKQKITDPEK
QTIEFGFRYLNAPQLYMNAWEKYNVSELDPRIKVTVKSFDDPRDNFNAYRFDG
VTGENLAWSADYNGVASTYDGLSGAALVPTLVKAHOTGDQKTKPEKAFIAKLA
KTLLEYAKTNKPNLSDFDEIDKTDNRVLNSNLPALIASYKLEKDPKGYAVIAKDV
KGSKPPIEFDFGKSSKSIETPDLYQWSAIFWASTFKNFTNEEISKIYEEFTFTV
LGYSDFVGRDKFKVLINKHYVSPYIAGSAVDSYQDNRILDKTKK"

3532..4677
/gene="oppb"
3532..4677
/gene="oppb"
/codon_start=1
/transl_table=4
/product="oligopeptide transport system permease protein
homolog"

/protein_id="CAA68074.1"
/db_xref="GI:1487983"
/db_xref="SPTREMBL:Q49552"
/translation="MTKYILORIAFATILFIISFVYVLVAFAFGNNPVNDLQAQOFQ
ASKSGLYTSLIKLEIENGRLYKNDPDLVPLIIVXGNYMKNFTKGDGFLYNPA
NNPSALYTNMAKLEFFLPLOXSLIITVPTFFISSILGIAIGVYAGYKRGKWFVSSW
FVLIFVAFSIFLIPVITISVALGISVVPVPRGEGSGFVNFVILPPIFVMSITSLA
GVYTRNQVITLTSNFVLIAKTGLSSSQIFEFKYLNRINISPLFLSFGSFGISLS
GSIIIEQWQVPGTSQIIVNAFTGCEINVMVFSTLFTFTFISLIADIIDVSYALDPK
IKYSTNSKRNWLFELAYRERSKLAKDLFKSLINKASQNTTVKGDN"

4679..5953
/gene="oppc"
4679..5953
/gene="oppc"
/codon_start=1
/transl_table=4
/product="oligopeptide transport system permease protein
homolog"

/protein_id="CAA68075.1"
/db_xref="GI:2911535"
/db_xref="SPTREMBL:Q49553"
/translation="MNINEVKAFNKEYKINTELEKFKFVAPEDRIVISSVAGPKPKL
GIEILKREFTKPVVILALAVFIILLISILVPANGISYVYKPNPDISDNFVKNLPAS
SPITVVALESNDPVPKFOYVLRNSETDPVKARMAFFLNGFOAKVNDVGVYAKYWR
VDLVVANYINGDIISKQITDAEYKLFNEOKNFYSHVLGTSIGYDITWATWATWR
AIKIALIVILEAIGISIGAFGLFYAGKLDITIMRVIEIFLAPPLIWLIFVSLI
GANESWLVTUTVGVPGFVGLTIRMFITTVKDEEFLIAAKAIGVGPFRFRVHALPTI
IGKIANNIVRSPVGVILVATLAFGLFGEKKKDTNLQMLIEASQEVGSNIWIHALPT
IILLFSLSLNFIALGVHDALDPRVMNKGK"

5955..7121
/gene="oppd"
5955..7121
/gene="oppd"
/codon_start=1
/transl_table=4

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

```

/product="oligopeptide transport ATP-binding protein
homolog"
/protein_id="CAA68076.1"
/db_xref="GI:2911536"
/db_xref="SPTREMBL:O53111"
/translation="MAKIINKHNDKNSSDTIHQKINKSTRVLDSIRVEPLKNNQI
LKRVNLHVFPOGHKSHIVRGVLDVHGEIIGLVGSGSGKSVTSKTLINNVESG
IISADQIOIGVEYELNKKQKWOVIRGQKIGYIPDPLTSLNPTNRVKGOLDALNN
NEDKNNSTTEKKNKILGLLTFEGLRSAEKIFYAPHILSGCMQKQRIYVVALKPE
LIADPTTALPTQVQSLAUFEDIRKMGISILILNSIVIAKFCDDIYVNIAGR
IVEKGTQKIEFTNPAHYTVALISAIPESKDEKLYTIGTTPDWMANLPLGDFPYRND
YALEIDYLKEPLPIITITHTAAATWLLSPEAPKIEKSKELOKRLDLFKKAFNE"
7114..9642
/gene="oppF"
CDS
7114..9642
/gene="oppF"
/codon_start=1
/transl_table=4
/product="oligopeptide transport ATP-binding protein
homolog"
/protein_id="CAA68077.1"
/db_xref="GI:2911537"
/db_xref="SPTREMBL:O53112"
/translation="MNNKFKDKKVLISIDMLKIFYVNOGMINKAVDGVSEFVHEGEI
VGLIGSGSKTIVGSTIDRLYDDYNGFVRLEDKIISGRISSESLNRLFRKNVOMIFQ
DPHASLNSQNIYSILKLEPLVNGYMKDKIKDFKDLNKKNKFYTFQEARMLEIQ
NYLEINKLAEPFLKWTENLNKLEFNDFISREDNFSFFGYLEEKQVSEIILNNMVA
NVDRLFEYFQRYRNKELPGQIALNEKEKLYIQCLSKTSKAYEASLELEKQ
KQIFNLRIKIEKISSEKNTFANYIHESKESLDIARLMSCDLDFSYNLKSELL
MKRGVLAKEIPKFKLGFNKKILINLEDEYNFVYKTLAPLPYSKTLKSDIKNV
LEKDFEFDVTKISELKQKLDSELQKLSREVEIKVYKKNPNEPTEISEQLOEAK
NLELQKQYLGSRKYLISYKAELESYLNKEQKQKYLLELRKQDQCNKYKELKQ
KFEEYVQURQEKILIAELQNDQKNDKSLAKQSKLSIKVQYNTILKM
YISDLSKEDTLKSDIERKYLDKQINNYLLGTDHKWETNLKTADAMAGVDKHM
KWRRKRDVSVPYIAKLISLITLYIKALEDVGLLKQFAYRPFEGGQK
RIVARALIVEQVIADEPIASDISIOAVNLLKELCKIKKNIGIFIAHDLSMTE
YVADEVIMHLCKIVESQKTEAIYANPIHPTINLFAIKPISNANEKFNQVSEALDY
LDEQKPNVPEFKVGEDHYVGTASQLKQWTHNAKLEKEHEL"
BASE COUNT 4157 a 1415 c 1501 g 3341 t
ORIGIN

Query Match          9.1%  Score 18;  DB 1;  Length 10414;
Best Local Similarity 100.0%;  Pred. No. 54;
Matches 18;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  53 AACATTAGGGAATTTTG 70
      |||||
Db  9534 AACATTAGGGAATTTTG 9517

RESULT 37
CEM18
LOCUS      CEM18      26542 bp      DNA      INV      02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid M18, complete sequence.
ACCESSION  Z68507
VERSION    Z68507.1  GI:1159944
KEYWORDS   HTG: Collagen; DNA repair protein like; Dyenin like.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans.
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 26542)
            Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
            Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
            Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
            Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
            Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
            Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
            O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopar, A.,
            Saunders, D., Showkheen, R., Smaldon, N., Smith, A., Sonhammer, E.,
            Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaubin, M.,
            Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
```

```

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Wilkinson-Sproat, J. and Wohldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
94150718
Steward, C.
Direct Submission
Submitted (16-JAN-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwen@matode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence&object=M18
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone M18. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone M18 is at 1 in this sequence. The true
right end of clone M18 is at 10009 in
sequence Z69634.
The true left end of clone B0001 is at 26439 in this sequence. The
true right end of clone F11A10 is at 23200 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
Z68297.
The end of this sequence (26439..26542) overlaps with the start of
sequence Z69634.
Location/Qualifiers
1..26542
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="M18"
5816..8308
/gene="M18.3"
join(5816..6232,7147..7315,7361..7494,7545..7993,
8050..8308)
/gene="M18.3"
/note="predicted using Genefinder: cDNA EST EMBL:D73035
comes from this gene; cDNA EST EMBL:D76008 comes from this
gene; cDNA EST YK587d1.3 comes from this gene"
/codon_start=1
/db_xref="GI:3878705"
/db_xref="SPTREMBL:O21555"
/translation="MGIDYNRWHLKDETIGEELKGYIKMIEEDLYVISTTSQLHVS
KLNSGSYNSIVNEILEQQGSNSIYYNQKLKLFIGRDVFGROSLVFNESHIFGCRIT
APFGSKWIEPFGQVTVFSLDSDPVIYSILDEYPENVLDDGIFSGSDISRIVYNNKH
GLNVSRTEIGEDISFNKDLQFKVIDATKILLKNYHSHVAVCLSGGVDSTFAHVW
HSSVPENQIDLINAVFGNSEKCEQAPDKRAKALESEFTAYPTQFRLILNVYDS
QTLVYNNKESIDAAQSPASVLDSDLSCLVFAVRAEVDSENOQSVRPATTCLLGS
GADELLAGYARHRTFRFEQITPENVAECENELRLGSRNGRDARVAQOLGTILSP
LLEDTVTLNALPVDSDKWLSPRGVGEKQLLRITVKMLGSPYDAPKQAMQFGSRMA
KMSNAGDNSIKGSKSPHLKTDYN"
complement(8729..10202)
/gene="M18.1"
complement(join(8729..9598,10059..10202))
/gene="M18.1"
/note="predicted using Genefinder: similar to collagen;
cDNA EST EMBL:D70719 comes from this gene; cDNA EST
EMBL:D66790 comes from this gene; cDNA EST EMBL:D65910
```

comes from this gene: cDNA EST EMBL:D65830 comes from this gene: cDNA EST EMBL:D66173 comes from this gene: cDNA EST EMBL:D69517 comes from this gene: cDNA EST EMBL:D70704 comes from this gene: cDNA EST EMBL:D65511 comes from this gene: cDNA EST EMBL:D70410 comes from this gene: cDNA EST EMBL:D68802 comes from this gene: cDNA EST EMBL:D66581 comes from this gene: cDNA EST EMBL:D66300 comes from this gene: cDNA EST EMBL:D68968 comes from this gene: cDNA EST EMBL:D69193 comes from this gene: cDNA EST EMBL:D69413 comes from this gene: cDNA EST EMBL:D69864 comes from this gene: cDNA EST EMBL:D70033 comes from this gene: cDNA EST EMBL:D70037 comes from this gene: cDNA EST EMBL:C10192 comes from this gene: cDNA EST EMBL:C11176 comes from this gene

/codon_start=1
/protein_id="CAA92826.1"
/db_xref="GI:3878706"
/db_xref="SPTREMBL:Q21556"
/translation="MSTARLVSIARAASTFAVLASIVSLVLDVNNYDDVTRNI ENFKDDANTVMHQMTVPPOGGKQQRQYVSLFGRKQRAQCNCAGNRAANCAGPP GPPAGDGDGHLVGEDGTPGADGNLGVPLSANGCIVCPAGPPGPGQONGGSG DQPGQDAMGGGQAGPGQAGDNGOPRPNDRGAPGAPGTRSGVRPGTAGSGP PQGPPGQNGDQPGDQAGPCQPCGHPQGNAGPPGTPGAAGQPGGDAAYCP PQSSRSVVRAQAARNRHVAARRHSVSRKAVARRSKARVVAAKHRVAKHRVAV OKA"

complement(111182..11457)
/gene="M18.2"
/complement(111182..11457)
/gene="M18.2"
/note="Similarity to C.reinhardtii dynein light chain (PIR Acc. No. A56444)"
/codon_start=1
/protein_id="CAA92827.1"
/db_xref="GI:3878707"
/db_xref="SWISS-PROT:Q21557"
/translation="MSEKIEVRETDMDPQDMVSVVREARQLYNIDKDVAAFKVE ELDKQATWIVICGKCFGRSVYEMGHFLLKCNVRGNQFIKCCA"
complement(15159..21081)
/gene="M18.5"
complement(join(15159..15220,15276..15381,15433..16249,16311..16437,16747..16970,17310..17644,18256..18957,19006..19330,19820..20021,20066..20354,20814..20956,21009..21081))
/gene="M18.5"
/note="predicted using GeneFinder; Similarity to C.aethiops (Green Monkey) possible DNA-repair protein XP-E (SW:XP_E.CERAE): cDNA EST EMBL:D27409 comes from this gene; cDNA EST EMBL:D27411 comes from this gene: cDNA EST EMBL:D27410 comes from this gene: cDNA EST EMBL:D34079 comes from this gene: cDNA EST EMBL:D34765 comes from this gene: cDNA EST EMBL:D32777 comes from this gene: cDNA EST EMBL:D35519 comes from this gene: cDNA EST EMBL:D37100 comes from this gene: cDNA EST yk224f3.5 comes from this gene: cDNA EST yk296b4.5 comes from this gene: cDNA EST yk324a2.5 comes from this gene: cDNA EST yk350c10.5 comes from this gene: cDNA EST yk430e3.5 comes from this gene"

/codon_start=1
/protein_id="CAA92824.2"
/db_xref="GI:5824558"
/db_xref="SPTREMBL:Q21554"
/translation="MPISYCVSARAKASVVSVEVNGTGHENVLNIVARGNIDVQLV SPEGLNKCEIPYIGVOLTIALVKCKRKRSLIVTEKWHMAILAIDRGKVVTRAAG CTAADTGRATNLFSLITHRNGLIAIRAFEGSVKMIOWESGTDLRHNFRFYPNVSD FFVFTGVDGNSVLYVYDDDHGKHLQFSLNMDHKEFRYSQASIAADSVLPVPP HAIGVIVGDSVLYVYKPNLDGVPVPTCSLNTFTTCHGIVDASGERLLSDTDG RLILMLNVTESQSYTVKEMRIDYLGTSIADSYINDGVVFGVSGRLSOLIRLM TPNNGSVVILETYSNIGPIRDMWVNESDQQLVCTGADKDGSLRVIRNGIGIDE LASVDLAGVGIFFPIRLSDNADNYVLSLDETHVLIQTIGELEDVKLEINIDLPIT FASTLFGPNDSGIILQATEKQIRLMSSSGSKSFWEPNTGEIISKVNAAANGQIVLAA RDTVYLITCIVDEMGAIDQLTAKMKFENETACDLSDNEGDNNKATFLVLAFWSTF AMEVQLPDLITVCTHPTKRIIPRSIATCIEEVHYLLVAFGDLVYVFDIKTGT HGEPKKSNGVTTPPSLHRVRNKRHLFCVSDRPVIFISAKKLIVSNVNVKIDVTVC SLSSSAYRDLIVSDGNSMVFCTVDIQIKHVRISIPNGESVLRVIAQKSTSTGVGCSN

RTSKARVFAASKNALVTSQSRPKVASTRADMDSEPPNTSSFMVLDQNTFVLHSHE FGPVETALSGISQFTNDSYVVGTLGYIDETETKIGRIVFEVDVDFERSKLRRV HELVVRGSLAIRLNGKLVAAINSIRLFEWTDKELRLECSFNFVIAIDLKVMRY EVAVADVRVSLSLYRMELGNEFEVAKQNSQMMVTCFETIAESILGGEAHLNFTV EYDTRITDDGRVLEPIGYWYLGELPKVMTRSTLVIQPEDSIIQYQSQIMFCTNOG TIGMIVQIDDKRKFLLAIKRAIDSVKNCMIEHSSTYTFVQRAARPPSGFVDDGL VESILDMDRSVAMDILSKVSDKGWPSLPDRPVEILKVIETDLARMH"

gene
join(21334..21559,21610..21835,21883..22236,22484..22724)
/gene="M18.6"
/gene="M18.6"
/codon_start=1
/protein_id="CAA92828.1"
/db_xref="GI:3878708"
/db_xref="SPTREMBL:Q21558"
/translation="MELIQDESESEDFYDAEDAEQAMEGLVNDVTHRFYALQDN ENRNYENEARLTDYTNWFPICGVSAVCFPQQLPMKFHDDEREIYERLVSARANG FVWLFPLDIPCTSLPYAIILPAHREKRSIATRINGYPNAIRDVTESAVNECVGLISG LPVDELPSLSPFIPKVAIHLPAHREKRSIATRINGYPNAIRDVTESAVNECVGLISG MLDGDTVROAKDGLTYESFAAKHIFSNRLTEYSLKEDSANGRIVAALKYAVIIIO CSRCHELHWEYISMTREPRRFFGIQREGITFQNELQEGDTEENWDHVPEDFDDQEE EDDAMNFRILILRR"

gene
join(23271..23478,23529..23648,23696..23787,24204..24330,24381..24556)
/gene="M18.7"
/gene="M18.7"
/note="similar to RNA recognition motif (aka RRM, RBD, or RNP domain); cDNA EST EMBL:T00930 comes from this gene; cDNA EST EMBL:T00219 comes from this gene; cDNA EST EMBL:T01089 comes from this gene; cDNA EST EMBL:T02416 comes from this gene; cDNA EST EMBL:D72966 comes from this gene; cDNA EST EMBL:D72973 comes from this gene; cDNA EST EMBL:D75932

Query Match 9.1%; Score 18; DB 34; Length 26542;
Best Local Similarity 100.08; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 TCACAGAAAAATATATA 103
|||||
Db 19709 TCACAGAAAAATATATA 19726

RESULT 38
CELLF46H5/c 38886 bp DNA INV 29-NOV-1996
LOCUS
DEFINITION
ACCESSION U41543
VERSION U41543.1 GI:1695234
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain-Bristol N2.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Peleoderinae; Caenorhabditis.
1 (bases 1 to 38886)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardiner, M., Kershaw, J., Kirsten, J., Laister, N., Johnston, L., Jones, M., Green, P., Hawkins, T., Hillier, L., Jier, M., Lathille, P., Lighthning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkneen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkison-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 38886)
AUTHORS Nhan, M.

TITLE The sequence of C. elegans cosmid F46H5
JOURNAL Unpublished (1995)
REFERENCE 3 (bases 1 to 3886)
AUTHORS Waterston.R.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-1995) Robert Waterston
REFERENCE 4 (bases 1 to 3886)
AUTHORS Waterston.R.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-1996)
COMMENT On Nov 29, 1996 this sequence version replaced gi:1109877.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwaterston@wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is ZK867, 200 bp overlap; 3' cosmid is T10A3, 800 bp
overlap. Actual start of this cosmid is at base position 1 of
CEL46H5; actual end is at 38286 of CEL46H5

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers

source

```
1. 38886
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F46H5"
/complement(225..1802)
/gene="F46H5.3"
complement(join(225..714,767..1092,1430..1636,1692..1802))
/gene="F46H5.3"
/Note="Similar to other ATP:guanine phosphotransferases;
coded for by C. elegans cDNA yk61c6.3; coded for by C.
elegans cDNA yk119h1.3; coded for by C. elegans cDNA
yk6a3.3; coded for by C. elegans cDNA CEMS843F; coded for
by C. elegans cDNA yk115b11.5; coded for by C. elegans
cDNA yk61c6.5; coded for by C. elegans cDNA cml4a3; coded
for by C. elegans cDNA CEMS144R; coded for by C. elegans
cDNA yk68a3.5; coded for by C. elegans cDNA yk82d10.5;
coded for by C. elegans cDNA CEMS86R; coded for by C.
elegans cDNA yk124f10.5; coded for by C. elegans cDNA
yk119h1.5; coded for by C. elegans cDNA CEMS86F; coded
for by C. elegans cDNA CEMS86Fb; coded for by C. elegans
cDNA yk142e5.3; coded for by C. elegans cDNA yk142e5.5;
coded for by C. elegans cDNA yk162g7.3; coded for by C.
elegans cDNA yk162g7.5; coded for by C. elegans cDNA
yk174e8.3; coded for by C. elegans cDNA yk174e8.5; coded
for by C. elegans cDNA yk184d9.3; coded for by C. elegans
cDNA yk184d9.5"
/codon_start=1
```

gene

CDS

```
/protein_id="BAB37022.1"
/db_xref="GI:1109878"
/translation="MLREAYSTSLHCVOAEGMSVSDVIAKIEBYAKLQAAPECHS
LLKYITREYVDQLKRTKILGLNDVIOGVNADGVYAPDAEATYTLFKPLFD
PLIOYHNGFAPDAQPNITDPEGLKFDPEKFINSTRIGRCSLOQYFPNCP
LSEANYLEMESKVAIFNDITDPELAGYFPLDGMTKEIQDLIKHDFLKGDGRLFO
AANACRYWPKGRGIFHNKOTFLIWCNEEDHLRIISNOEGNGVQVLEIRLKGVTIE
KOAPFSRDDRLGWLTFPCSNLGTVRASVHIRLPKISAKPDFKSIDGLKLQIRGIG
EHSEEGCVYDISNKAELGTEFEAVQMYDGIHAHLALEKAA"
complement(8345..12126)
/gene="F46H5.7"
complement(join(8345..8450,8498..8586,8630..8789,
8840..8920,9422..9643,9693..9914,9964..10083,10136..10297,
10356..10572,11139..11231,11324..11397,11884..12022,
12093..12126))
/gene="F46H5.7"
/Note="Coded for by C. elegans cDNA yk100c6.3; coded for
by C. elegans cDNA yk100c6.5; coded for by C. elegans cDNA
yk134h9.5; coded for by C. elegans cDNA cm08f2; coded for
by C. elegans cDNA cm20g12"
/codon_start=1
/protein_id="BAB37024.1"
/db_xref="GI:1109879"
/translation="MADVATKIPDANVRLEIEPVQVSHDQDATLNFATVQSVGCAH
GLYVRDGGKALLFDNNTGKVPYPPNGWDSVPITFHLAHSRHGTHFADYSKANDAF
KNISAVOKLFAAGLGRGSOQVYVAACTRASTSKKNQPSQFDAQPNVNISESE
GRDSEKENDQLEKLRNDELAKIRVEYKSDLSAQLKXLEKLESQASQSESE
KFRVDSIDKRSQDDEEINLHNHVEELRNKLATIVESKNDLCTKSEQDYLKN
ADKVTFLNQINSDAHEEVASTAVRALEVLNGLANNIRQAEEAKQQLQEANVANE
RVGKLEQNGYLGITEQLKARADTSHAERMLKDKSEKRVVEINEESKLEIRLGEUSQ
WMDAKRWGELESVALQRLDNTANSKIQSLNDQSHSSTMTPTDGTIISOSNOGT
WNLANGNPOCTGAIPISPLVSALPAGANPFIFGANEGRNVTITNTIOEYVLT
GSFINWKTLCCKELVSGKGVTVNLTGRGHEFRFINGEWATSSDYQQVNGLGGQN
NIIFVE"
complement(14754..23897)
/gene="F46H5.4"
complement(join(14754..15029,15678..15873,15926..16021,
16070..16299,16348..16711,16766..16981,17030..17235,
17286..17447,17491..17673,17891..18148,18201..18359,
18406..18703,18752..18871,18925..19098,19147..19418,
19465..19610,19655..19735,19899..20133,20184..20474,
20522..20660,20711..20820,20865..21061,21113..21316,
21360..21521,21567..21716,21764..21884,21940..22000,
22060..22214,22261..22448,22513..22706,23287..23352,
23400..23486,23535..23638,23684..23792,23851..23897))
/gene="F46H5.4"
/Note="Similar to Rat trig gene product; coded for by C.
elegans cDNA yk31e7.5; coded for by C. elegans cDNA
yk40d6.5; coded for by C. elegans cDNA yk31e7.3; coded for
by C. elegans cDNA yk40d6.3; coded for by C. elegans cDNA
yk149g5.3; coded for by C. elegans cDNA yk149g5.5; coded
for by C. elegans cDNA yk153h2.3; coded for by C. elegans
cDNA yk153h2.5"
/codon_start=1
/protein_id="BAB37023.1"
/db_xref="GI:1109880"
/translation="MKPADGRARFVKRSKVTASEVRKXHVISGLHIHRLSGDNLGLD
LAVSMKEIQETPEVPVVDVALLDQRKCLLYSISPKRSKFLFVSDDEVITIR
ODGLDHTYTTIPNIENVRDIDCGYDNFSLNRYAQLGTEIDKRNLEKLTALR
SLRPIQFINGIRIMDRPSTISIDGSDFTDPTKCAEQEYVYSMLKSAIDKNTLLY
SMASNRVYKFMNCINEEKSWMKRGVPLPEQETPKLVYKKAADAPFPEFLF
SMALYDINKRQVTESYFYNIADHDKLMDLGSHPNFNNYMQVLFNVTGLKEDMFL
VYIKVQLQNDVENSEPYTGTCDENNERLERAEKNCORLGAVRSGPLGFOVDLO
RIYKNAVSTGASSDRDTPMMSQCTASCAYLTAGOSQDDOCSITENDRTSIASM
GSTLRFSGSQAATVFSRVRTPLTKKFPVSNLPTSQVPIENINPSCNLKFSF
IKQEGDKTSSDDIRIGENRRKTKVHKFMNFELLETLAGNSKSKREYSHGSNULTL
NSERVTHAMEIPIYQASLNKYNKIVFYPRKINLSNRTGNARNMTKIELMDANET
AOEVFFGSTRMFLTSKTSYIHNRTPHFTDEIKLSDPLCDLNDHLLFTVYHS
KEGSSSTESPISGYTWLPLXRNKLRSGNHFPLVCEKPPRPGYGLDNNALPNLKW
VNHKPIFSCSTEVISVHAQDEFLFNLAGVASLSDNPKPVPQETQILIRSEGLC
KTEPKLTAFTFTIMSRLFLTIAPPYSDLSMKAFYICGLKLFNSVHLDLDAHQ
RMLLVSPYVKYKQAAQESKPSNIRPVELKSSPTONSLSISMEIHERTHSAVNTGT
KNIRHECLLEVLWRGSLRDSVLVHSWFLLEILLKSCSEYLTMTGRHSPRKSREE
```

gene

CDS

gene

CDS

EQFLKNNLTUVDILAQEVIRHNDPKARMIISNSLGVFLRDCFSIMDRFTVMKLVHK
YLIAFAESMRKLVHSNELLISIKIDFVYVVCYSYEHLYLVNLSLDLKPSTNGGAVPPA
SFLSGNRTKSSSTWTLNDISRSTHYLSQGVLSMDKDSIISGKTLLCAKAIETVKEL
LOSHDDDRIVEGESLAOVANIYKPLVTIVLONIECLHSGSVRNSTDYSSNSTSVFEOT
ORDDVAAIAGKLRNSPDPTFGKQOMDLPWTKTILCCVFWVLNRDIDREDLKHWRSLD
NENKMLHLHFHTMTSFEIKDDPASARSPKTSILTKLDEPEPGQVYKRAQSETC
ETKVQSTIDUALSSAIVSCEVFMCVIEVDNIITAVATDPRNAOFHILPFIPLIMH
GLSCNADQVLEVFIAAQONFLAKFPDPMILEONPELCAELSQOILRHCSSTRELVRT
MAVSYLHFURENFKLYRNLTRARTLSTALSTLSSGCGIDIFYNDEFMRSLERIAN
QLAADDITDIAAKKLTQOMQELTANLQIMLSTVREHVNDYEMTIDLMYOLVEG
YSNPDRLTWLLNMAERHEKQKNLCEAAHSYLOASALVFETIAOKDONLSPESKGA
TFSEITPNALKESTNFNSLNKADSENHIQSYHFTAGVIKILERAFALLKKAOLYEL
LFPESTILKAYNTKYSRSVSTHRELGIADOLKETGEYVENQSDAWISPLGIDK
RCGSPFRVAFYKALGALNAEYVYKESAFKLSNEISNRLETFITMYGAGNVVVLK
DAPVQLEKFNPKAKIOITFVDVLSNKEKERTYFTFRNNVNRFFYEAPYTMGR
AQGLAAQYKKRILTVENSFPYIKTRLVNRSVKDFSPIEVAIEDIEKKTRELSAA
AOKNPKMLSMFTQGSIGTIVNOGPLEIANVFLANMLDDRGPRVDRLOKRLRSFRH
LQCKAMEAIELQLIGEDQKEYORNVENEFSEFVTHLKPMLSRQNETITSEFSIPT
VV"

gene complement(25261..26312)
/gene="F46H5.5"
CDS complement(join(25261..25392,25442..25581,25632..25722,
25768..26131,26185..26312))
/gene="F46H5.5"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB37026.1"

Query Match 9.18; Score 18; DB 34; Length 38886;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 CACTAGAAAATATATAT 104
|||||

Db 7478 CACTAGAAAATATATAT 7461

RESULT 39
AC008437/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_329A22, *** SEQUENCING IN
PROGRESS ***, 31 unordered pieces.
ACCESSION AC008437
VERSION AC008437.1 GI:5686629
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 41402)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 41402)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 843: contig of 843 bp in length
* gap of unknown length
* 844 1684: contig of 841 bp in length
* gap of unknown length
* 1685 2545: contig of 861 bp in length
* gap of unknown length

* 2546 3353: contig of 808 bp in length
* gap of unknown length
* 3354 4230: contig of 877 bp in length
* gap of unknown length
* 4231 5091: contig of 861 bp in length
* gap of unknown length
* 5092 6040: contig of 949 bp in length
* gap of unknown length
* 6041 7557: contig of 1517 bp in length
* gap of unknown length
* 7558 8758: contig of 1201 bp in length
* gap of unknown length
* 8759 9648: contig of 890 bp in length
* gap of unknown length
* 9649 10732: contig of 1084 bp in length
* gap of unknown length
* 10733 11784: contig of 1052 bp in length
* gap of unknown length
* 11785 12650: contig of 866 bp in length
* gap of unknown length
* 12651 13529: contig of 879 bp in length
* gap of unknown length
* 13530 14784: contig of 1255 bp in length
* gap of unknown length
* 14785 16385: contig of 1601 bp in length
* gap of unknown length
* 16386 17197: contig of 812 bp in length
* gap of unknown length
* 17198 18040: contig of 843 bp in length
* gap of unknown length
* 18041 19499: contig of 1459 bp in length
* gap of unknown length
* 19500 20458: contig of 959 bp in length
* gap of unknown length
* 20459 21887: contig of 1429 bp in length
* gap of unknown length
* 21888 23384: contig of 1477 bp in length
* gap of unknown length
* 23365 25150: contig of 1786 bp in length
* gap of unknown length
* 25151 27026: contig of 1876 bp in length
* gap of unknown length
* 27027 28817: contig of 1791 bp in length
* gap of unknown length
* 28818 30556: contig of 1739 bp in length
* gap of unknown length
* 30557 32533: contig of 1977 bp in length
* gap of unknown length
* 32534 34180: contig of 1647 bp in length
* gap of unknown length
* 34181 36262: contig of 2082 bp in length
* gap of unknown length
* 36263 38508: contig of 2246 bp in length
* gap of unknown length
* 38509 41402: contig of 2894 bp in length.
FEATURES
Location/Qualifiers
source 1..41402
organism="Homo sapiens"
db_xref="taxon:9606"
chromosome="5"
clone="CIT-HSPC_329A22"

BASE COUNT 12577 a 7525 c 7689 g 13482 t 129 others
ORIGIN

Query Match 9.18; Score 18; DB 41; Length 41402;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 118 ATAATTTTGAGATAATTA 135
|||||

Db 16210 ATAATTTTGAGATAATTA 16193

```

RESULT 40
AC016648/c
LOCUS
DEFINITION
AC016648 45617 bp DNA HTG 04-DEC-1999
Homo sapiens chromosome 5 clone RPCI-1_137K24, *** SEQUENCING IN
PROGRESS ***, 40 unordered pieces.
ACCESSION
AC016648
VERSION
AC016648.1 GI:5524058
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 45617)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1042: contig of 1042 bp in length
1 gap of unknown length
* 1043 1777: contig of 735 bp in length
* gap of unknown length
* 1778 2698: contig of 921 bp in length
* gap of unknown length
* 2699 3417: contig of 719 bp in length
* gap of unknown length
* 3418 4250: contig of 833 bp in length
* gap of unknown length
* 4251 4988: contig of 738 bp in length
* gap of unknown length
* 4989 6185: contig of 1197 bp in length
* gap of unknown length
* 6186 7030: contig of 845 bp in length
* gap of unknown length
* 7031 7964: contig of 934 bp in length
* gap of unknown length
* 7965 8741: contig of 776 bp in length
* gap of unknown length
* 8741 9732: contig of 992 bp in length
* gap of unknown length
* 9733 10862: contig of 1130 bp in length
* gap of unknown length
* 10863 11583: contig of 721 bp in length
* gap of unknown length
* 11584 12502: contig of 919 bp in length
* gap of unknown length
* 12503 13207: contig of 705 bp in length
* gap of unknown length
* 13208 14250: contig of 1043 bp in length
* gap of unknown length
* 14251 15560: contig of 1310 bp in length
* gap of unknown length
* 15561 16481: contig of 921 bp in length
* gap of unknown length
* 16482 17643: contig of 1162 bp in length
* gap of unknown length
* 17644 18527: contig of 884 bp in length
* gap of unknown length
* 18528 19624: contig of 1097 bp in length
* gap of unknown length
* 19625 20712: contig of 1088 bp in length

```

```

* 20713 21946: contig of 1234 bp in length
* gap of unknown length
* 21947 23095: contig of 1149 bp in length
* gap of unknown length
* 23096 24350: contig of 1255 bp in length
* gap of unknown length
* 24351 25672: contig of 1322 bp in length
* gap of unknown length
* 25673 26629: contig of 957 bp in length
* gap of unknown length
* 26630 27703: contig of 1074 bp in length
* gap of unknown length
* 27704 28979: contig of 1276 bp in length
* gap of unknown length
* 28980 30321: contig of 1342 bp in length
* gap of unknown length
* 30322 31485: contig of 1164 bp in length
* gap of unknown length
* 31486 32601: contig of 1116 bp in length
* gap of unknown length
* 32602 34106: contig of 1505 bp in length
* gap of unknown length
* 34107 35775: contig of 1669 bp in length
* gap of unknown length
* 35776 37641: contig of 1866 bp in length
* gap of unknown length
* 37642 39269: contig of 1628 bp in length
* gap of unknown length
* 39270 40997: contig of 1728 bp in length
* gap of unknown length
* 40998 42877: contig of 1880 bp in length
* gap of unknown length
* 42878 43958: contig of 1081 bp in length
* gap of unknown length
* 43959 45617: contig of 1659 bp in length.
FEATURES
    source
    1..45617
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RPCI-1_137K24"
BASE COUNT 13659 a 8800 c 8954 g 14163 t 41 others
ORIGIN

Query Match          9.1%; Score 18; DB 44; Length 45617;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 AAGAACTCTATTATATGA 195
      |||||
Db 28111 AAGAACTCTATTATATGA 28094

RESULT 41
AC012518/c
LOCUS
DEFINITION
AC012518 57784 bp DNA HTG 30-OCT-1999
Homo sapiens clone RP11-512G13, *** SEQUENCING IN PROGRESS ***, 48
unordered pieces.
ACCESSION
AC012518
VERSION
AC012518.1 GI:6139063
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 57784)
AUTHORS
Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunag,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,X., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,

```

Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M., Holloway, C., Hosak, H., Jackson, I.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondajewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Osval, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

Direct Submission
2 (bases 1 to 57784)
Worley, K.C.
Direct Submission
Submitted (29-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 2443: contig of 2443 bp in length
* 2444
* 2494: gap of unknown length
* 2494
* 4658: contig of 2165 bp in length
* 4659
* 4708: gap of unknown length
* 4709
* 6857: contig of 2149 bp in length
* 6858
* 6907: gap of unknown length
* 6908
* 8900: contig of 1993 bp in length
* 8901
* 8950: gap of unknown length
* 8951
* 10881: contig of 1931 bp in length
* 10882
* 10931: gap of unknown length
* 10932
* 12724: contig of 1793 bp in length
* 12725
* 12774: gap of unknown length
* 12775
* 14539: contig of 1765 bp in length
* 14540
* 14586: gap of unknown length
* 14590
* 16298: contig of 1709 bp in length
* 16299
* 16348: gap of unknown length
* 16349
* 17854: contig of 1506 bp in length
* 17855
* 17904: gap of unknown length
* 17905
* 19409: contig of 1505 bp in length
* 19410
* 19459: gap of unknown length
* 19460
* 20904: contig of 1445 bp in length
* 20905
* 20954: gap of unknown length
* 20955
* 22380: contig of 1426 bp in length
* 22381
* 22430: gap of unknown length
* 22431
* 23849: contig of 1419 bp in length
* 23850
* 23899: gap of unknown length
* 23900
* 25287: contig of 1388 bp in length
* 25288
* 25337: gap of unknown length
* 25338
* 26711: contig of 1374 bp in length
* 26712
* 26761: gap of unknown length
* 26762
* 28117: contig of 1356 bp in length
* 28118
* 28167: gap of unknown length
* 28168
* 29457: contig of 1290 bp in length
* 29458
* 29507: gap of unknown length
* 29508
* 30793: contig of 1286 bp in length
* 30794
* 30843: gap of unknown length
* 30844
* 32113: contig of 1270 bp in length
* 32114
* 32163: gap of unknown length
* 32164
* 33419: contig of 1256 bp in length
* 33420
* 33469: gap of unknown length
* 33470
* 34675: contig of 1206 bp in length
* 34676
* 34725: gap of unknown length
* 35930: contig of 1205 bp in length
* 35931
* 35980: gap of unknown length

35981
37186
37186
37335: gap of unknown length
37335: contig of 1205 bp in length
38362: contig of 1127 bp in length
38412: gap of unknown length
38413
39535: contig of 1123 bp in length
39585: gap of unknown length
39585: gap of unknown length
40589: contig of 1074 bp in length
40590
40709: gap of unknown length
41623: contig of 914 bp in length
41673: gap of unknown length
42528: contig of 855 bp in length
42578: gap of unknown length
43427: contig of 849 bp in length
43477: gap of unknown length
43478
43722: contig of 845 bp in length
43722: gap of unknown length
43771: gap of unknown length
45187: contig of 816 bp in length
45236: gap of unknown length
46042: contig of 806 bp in length
46091: gap of unknown length
46892: contig of 801 bp in length
46941: gap of unknown length
47740: contig of 799 bp in length
47789: gap of unknown length
48120: contig of 331 bp in length
48129: gap of unknown length
48169: gap of unknown length
48504: contig of 335 bp in length
48553: gap of unknown length
48554
49348: contig of 795 bp in length
49397: gap of unknown length
50192: contig of 795 bp in length
50241: gap of unknown length
51036: contig of 795 bp in length
51085: gap of unknown length
51085: gap of unknown length
51877: contig of 792 bp in length
51926: gap of unknown length
52171: contig of 791 bp in length
52766: gap of unknown length
53557: contig of 791 bp in length
53606: gap of unknown length
54395: contig of 789 bp in length
54444: gap of unknown length
55232: contig of 788 bp in length
55281: gap of unknown length
55598: contig of 317 bp in length
55647: gap of unknown length
55599
56426: contig of 779 bp in length
56475: gap of unknown length
57175: contig of 700 bp in length
57224: gap of unknown length
57784: contig of 560 bp in length.

FEATURES
source

Location/Qualifiers
1..57784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-512G13"

BASE COUNT 16242 a 11269 c 11244 g 16579 t 2450 others
ORIGIN

Query Match 9.1%; Score 18; DB 42; Length 57784;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 AAAATATATATATGGTG 111
|||||
Db 32194 AAAATATATATATGGTG 32177

RESULT 42
AC010280/c

AC010280 70390 bp DNA HTG 15-SEP-1999
LOCUS Homo sapiens chromosome 5 clone CIT-HSPC_537E7, *** SEQUENCING IN

ACCESSION AC010280
VERSION AC010280.1 GI:5882653
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 70390)
AUTHORS DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
TITLE Unpublished
REFERENCE 2 (bases 1 to 70390)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA
www.jgi.doe.gov.
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 84 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 596: contig of 596 bp in length
gap of unknown length
597 1213: contig of 617 bp in length
gap of unknown length
1214 1865: contig of 652 bp in length
gap of unknown length
1866 2428: contig of 563 bp in length
gap of unknown length
2429 3529: contig of 1101 bp in length
gap of unknown length
3530 4154: contig of 625 bp in length
gap of unknown length
4155 4933: contig of 779 bp in length
gap of unknown length
4934 5574: contig of 641 bp in length
gap of unknown length
5575 6224: contig of 650 bp in length
gap of unknown length
6225 6962: contig of 738 bp in length
gap of unknown length
6963 7608: contig of 646 bp in length
gap of unknown length
7609 8291: contig of 683 bp in length
gap of unknown length
8292 8900: contig of 609 bp in length
gap of unknown length
8901 9555: contig of 655 bp in length
gap of unknown length
9556 10191: contig of 636 bp in length
gap of unknown length
10192 10905: contig of 714 bp in length
gap of unknown length
10906 11605: contig of 700 bp in length
gap of unknown length
11606 12318: contig of 713 bp in length
gap of unknown length
12319 12938: contig of 620 bp in length
gap of unknown length
12939 13608: contig of 670 bp in length
gap of unknown length
13609 14273: contig of 665 bp in length
gap of unknown length
14274 14850: contig of 577 bp in length
gap of unknown length
14851 15920: contig of 1070 bp in length
gap of unknown length
15921 16601: contig of 681 bp in length
gap of unknown length
16602 17878: contig of 1277 bp in length
gap of unknown length
17879 18931: contig of 1053 bp in length
gap of unknown length
18932 19586: contig of 655 bp in length
gap of unknown length
19587 20254: contig of 668 bp in length
gap of unknown length
20255 21173: contig of 919 bp in length
gap of unknown length
21174 21856: contig of 683 bp in length
gap of unknown length
21857 22476: contig of 620 bp in length
gap of unknown length
22477 23307: contig of 831 bp in length
gap of unknown length
23308 24000: contig of 693 bp in length
gap of unknown length
24001 24728: contig of 728 bp in length
gap of unknown length
24729 25396: contig of 668 bp in length
gap of unknown length
25397 26092: contig of 696 bp in length
gap of unknown length
26093 26761: contig of 669 bp in length
gap of unknown length
26762 28159: contig of 1398 bp in length
gap of unknown length
28160 28839: contig of 680 bp in length
gap of unknown length
28840 29509: contig of 670 bp in length
gap of unknown length
29510 30209: contig of 700 bp in length
gap of unknown length
30210 30932: contig of 723 bp in length
gap of unknown length
30933 31588: contig of 656 bp in length
gap of unknown length
31589 32225: contig of 637 bp in length
gap of unknown length
32226 32854: contig of 629 bp in length
gap of unknown length
32855 33476: contig of 622 bp in length
gap of unknown length
33477 34076: contig of 600 bp in length
gap of unknown length
34077 34812: contig of 736 bp in length
gap of unknown length
34813 35457: contig of 645 bp in length
gap of unknown length
35458 36131: contig of 674 bp in length
gap of unknown length
36132 36774: contig of 643 bp in length
gap of unknown length
36775 37369: contig of 595 bp in length
gap of unknown length
37370 38026: contig of 657 bp in length
gap of unknown length
38027 38662: contig of 636 bp in length
gap of unknown length
38663 39286: contig of 624 bp in length
gap of unknown length
39287 39370: contig of 84 bp in length
gap of unknown length
39371 40404: contig of 1034 bp in length
gap of unknown length
40405 41437: contig of 1033 bp in length
gap of unknown length
41438 42318: contig of 881 bp in length
gap of unknown length
42319 42486: contig of 168 bp in length
gap of unknown length

* 42487 43151: contig of 665 bp in length
 * gap of unknown length
 * 43152 44251: contig of 1100 bp in length
 * gap of unknown length
 * 44252 45291: contig of 1040 bp in length
 * gap of unknown length
 * 45292 46221: contig of 930 bp in length
 * gap of unknown length
 * 46222 47473: contig of 1252 bp in length
 * gap of unknown length
 * 47474 48805: contig of 1332 bp in length
 * gap of unknown length
 * 48806 49527: contig of 722 bp in length
 * gap of unknown length
 * 49528 50567: contig of 1040 bp in length
 * gap of unknown length
 * 50568 52219: contig of 1652 bp in length
 * gap of unknown length
 * 52220 53518: contig of 1299 bp in length
 * gap of unknown length
 * 53519 53635: contig of 117 bp in length
 * gap of unknown length
 * 53636 54472: contig of 837 bp in length
 * gap of unknown length
 * 54473 55286: contig of 814 bp in length
 * gap of unknown length
 * 55287 56876: contig of 1590 bp in length
 * gap of unknown length
 * 56877 57989: contig of 1113 bp in length
 * gap of unknown length
 * 57990 59011: contig of 1022 bp in length
 * gap of unknown length
 * 59012 60361: contig of 1350 bp in length
 * gap of unknown length
 * 60362 61992: contig of 1631 bp in length
 * gap of unknown length
 * 61993 63418: contig of 1426 bp in length
 * gap of unknown length
 * 63419 64754: contig of 1336 bp in length
 * gap of unknown length
 * 64755 65992: contig of 1238 bp in length
 * gap of unknown length
 * 65993 67163: contig of 1171 bp in length
 * gap of unknown length
 * 67164 68719: contig of 1556 bp in length
 * gap of unknown length
 * 68720 70390: contig of 1671 bp in length.

FEATURES

source
 1. .70390
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"

Query Match 9.1%; Score 18; DB 41; Length 70390;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TAGAAAAATATATATT 107
 |||||
 Db 58294 TAGAAAAATATATATT 58277

RESULT 43

AB024028 70952 bp DNA PLN 20-NOV-1999
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K1G2,
 complete sequence.
 AB024028
 ACCESSION
 AB024028.1 GI:4519187
 VERSION
 HTG
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
 Clone:K1G2.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.

REFERENCE

AUTHORS Nakamura,Y.
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. I
 JOURNAL Unpublished (1998)
 REFERENCE 2 (bases 1 to 70952)

AUTHORS

Nakamura,Y.

TITLE

Direct Submission

JOURNAL

Submitted (24-FEB-1999) to the DDBJ/EMBL/GenBank databases.

Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of

Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan

(E-mail:ynakamu@kazusa.or.jp, tel:+81-438-52-3935,

Fax:+81-438-52-3934)

FEATURES

Location/Qualifiers
 source
 1. .70952
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone_lib="Mitsui P1"
 /clone="K1G2"

BASE COUNT 23075 a 12412 c 12631 g 22834 t

ORIGIN

Query Match 9.1%; Score 18; DB 7; Length 70952;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 AGAAAAATATATATTG 108
 |||||

Db 16661 AGAAAAATATATATTG 16678

RESULT 44

AC016217 71293 bp DNA HTG 23-NOV-1999
 LOCUS Homo sapiens clone RP11-26F18, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION
 AC016217
 ACCESSION
 AC016217.1 GI:6466577
 VERSION
 HTG: HTGS_PHASE0.
 KEYWORDS
 human.

SOURCE

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 1 (bases 1 to 71293)

REFERENCE

AUTHORS

TITLE

Homo sapiens, clone RP11-26F18

JOURNAL

Unpublished

AUTHORS

2 (bases 1 to 71293)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Bozulavkiy,L., Boukhgaiter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karacas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

```
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: 14628
Center clone name: 26_F_18
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 845: contig of 845 bp in length
* gap of unknown length
* 846 1716: contig of 871 bp in length
* gap of unknown length
* 1717 2589: contig of 873 bp in length
* gap of unknown length
* 2590 3446: contig of 857 bp in length
* gap of unknown length
* 3447 4294: contig of 848 bp in length
* gap of unknown length
* 4295 5182: contig of 888 bp in length
* gap of unknown length
* 5183 6077: contig of 895 bp in length
* gap of unknown length
* 6078 6965: contig of 888 bp in length
* gap of unknown length
* 6966 7812: contig of 847 bp in length
* gap of unknown length
* 7813 8707: contig of 895 bp in length
* gap of unknown length
* 8708 9587: contig of 880 bp in length
* gap of unknown length
* 9588 10407: contig of 820 bp in length
* gap of unknown length
* 10408 11312: contig of 905 bp in length
* gap of unknown length
* 11313 12178: contig of 866 bp in length
* gap of unknown length
* 12179 13063: contig of 885 bp in length
* gap of unknown length
* 13064 13929: contig of 866 bp in length
* gap of unknown length
* 13930 14804: contig of 875 bp in length
* gap of unknown length
* 14805 15681: contig of 877 bp in length
* gap of unknown length
* 15682 16549: contig of 868 bp in length
* gap of unknown length
* 16550 17445: contig of 896 bp in length
* gap of unknown length
* 17446 18279: contig of 834 bp in length
* gap of unknown length
* 18280 19157: contig of 878 bp in length
* gap of unknown length
* 19158 20016: contig of 859 bp in length
* gap of unknown length
* 20017 20861: contig of 845 bp in length
* gap of unknown length
* 20862 21726: contig of 865 bp in length
* gap of unknown length
* 21727 22606: contig of 880 bp in length
* gap of unknown length
* 22607 23518: contig of 912 bp in length
* gap of unknown length
*
* 23519 24371: contig of 853 bp in length
* gap of unknown length
* 24372 25255: contig of 884 bp in length
* gap of unknown length
* 25256 26130: contig of 875 bp in length
* gap of unknown length
* 26131 27011: contig of 881 bp in length
* gap of unknown length
* 27012 27910: contig of 899 bp in length
* gap of unknown length
* 27911 28780: contig of 870 bp in length
* gap of unknown length
* 28781 29666: contig of 886 bp in length
* gap of unknown length
* 29667 30535: contig of 869 bp in length
* gap of unknown length
* 30536 31418: contig of 883 bp in length
* gap of unknown length
* 31419 32331: contig of 913 bp in length
* gap of unknown length
* 32332 33232: contig of 901 bp in length
* gap of unknown length
* 33233 34112: contig of 880 bp in length
* gap of unknown length
* 34113 35013: contig of 901 bp in length
* gap of unknown length
* 35014 35911: contig of 898 bp in length
* gap of unknown length
* 35912 36814: contig of 903 bp in length
* gap of unknown length
* 36815 37703: contig of 889 bp in length
* gap of unknown length
* 37704 38555: contig of 852 bp in length
* gap of unknown length
* 38556 39457: contig of 902 bp in length
* gap of unknown length
* 39458 40330: contig of 873 bp in length
* gap of unknown length
* 40331 41212: contig of 882 bp in length
* gap of unknown length
* 41213 42096: contig of 884 bp in length
* gap of unknown length
* 42097 42972: contig of 876 bp in length
* gap of unknown length
* 42973 43868: contig of 896 bp in length
* gap of unknown length
* 43869 44749: contig of 881 bp in length
* gap of unknown length
* 44750 45637: contig of 888 bp in length
* gap of unknown length
* 45638 46504: contig of 867 bp in length
* gap of unknown length
* 46505 47375: contig of 871 bp in length
* gap of unknown length
* 47376 48265: contig of 890 bp in length
* gap of unknown length
* 48266 49145: contig of 880 bp in length
* gap of unknown length
* 49146 50036: contig of 881 bp in length
* gap of unknown length
* 50037 50906: contig of 880 bp in length
* gap of unknown length
* 50907 51796: contig of 890 bp in length
* gap of unknown length
* 51797 52661: contig of 865 bp in length
* gap of unknown length
* 52662 53549: contig of 888 bp in length
* gap of unknown length
* 53550 54437: contig of 888 bp in length
* gap of unknown length
* 54438 55331: contig of 894 bp in length
* gap of unknown length
```

```
* 55332 56211: contig of 880 bp in length
* 56212 57077: contig of 866 bp in length
* 57078 57963: contig of 886 bp in length
* 57964 58866: contig of 903 bp in length
* 58867 59738: contig of 872 bp in length
* 59739 60635: contig of 897 bp in length
* 60636 61519: contig of 884 bp in length
* 61520 62425: contig of 906 bp in length
* 62426 63299: contig of 874 bp in length
* 63300 64178: contig of 879 bp in length
*          gap of unknown length

Query Match          9.1% Score 18; DB 44; Length 71293;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AGCATTTCACAGACAAG 166
|||||
Db 34863 AGCATTTCACAGACAAG 34880

RESULT 45
AF191071/c          88481 bp      DNA      PRI      11-OCT-1999
LOCUS              Homo sapiens chromosome 8 clone BAC 388D06, complete sequence.
DEFINITION          AF191071
ACCESSION            AF191071.1 GI:6018318
VERSION              HTG.
KEYWORDS              human.
SOURCE               ORGANISM
                     Homo sapiens
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                     Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (bases 1 to 88481)
AUTHORS              Reichwald,K., Baumgart,C., Dette,M., Menzel,U., Schillhabel,M.,
                     Weng,G. and Rosenthal,A.
TITLE                Direct Submission
JOURNAL              Submitted (29-SEP-1999) Genome Analysis, Institute of Molecular
                     Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
                     source
                     1..88481
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="BAC 388D06"
                     1..72
                     /rpt_family="L1P3_5end"
repeat_region        68..1499
repeat_region        286..563
exon                 /note="GRAIL"
                     /evidence=not_experimental
repeat_region        1353..2244
exon                 /rpt_family="L1PA8"
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        2256..2564
repeat_region        2789..2811
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
repeat_region        2880..2932
                     /note="AT_rich"
                     /rpt_family="Low_complexity"

3366..3388
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
4271..4750
/note="MZE"
/evidence=not_experimental
complement(4378..4937)
/note="GENSCAN"
complement(4397..4422)
/note="XPOUND"
/evidence=not_experimental
4468..4490
/note="XPOUND"
/evidence=not_experimental
4492..4514
/note="XPOUND"
/evidence=not_experimental
complement(5171..5319)
/rpt_family="MIR"
5471..5638
/note="MZE"
/evidence=not_experimental
complement(5859..5901)
/note="AT_rich"
/rpt_family="Low_complexity"
5975..6002
/note="AT_rich"
/rpt_family="Low_complexity"
complement(7160..7447)
/rpt_family="AluX"
complement(7891..7933)
/note="(TAGA)n"
/rpt_family="Simple_repeat"
8060..8121
/rpt_family="MLT1"
complement(8823..8905)
/note="MZE"
/evidence=not_experimental
9007..9100
/note="AT_rich"
/rpt_family="Low_complexity"
complement(9632..10675)
/rpt_family="L1MCI"
10202..10334
/note="MZE"
/evidence=not_experimental
complement(10543..13163)
/rpt_family="L1M4_orf2"
complement(13170..13373)
/rpt_family="MLT2B"
complement(13319..13358)
/evidence=not_experimental
13374..14061
/rpt_family="L1ME1"
14070..14283
/rpt_family="L1MEC_5end"
14177..14234
/note="MZE"
/evidence=not_experimental
15313..15404
/note="GRAIL"
/evidence=not_experimental
15313..15404
/note="MZE"
/evidence=not_experimental
complement(16097..16540)
/rpt_family="L1MEa_5end"
16560..18253
/rpt_family="L1M2_orf2"
complement(17617..17816)
/rpt_family="MER30"
complement(17635..17698)
```

```
/note="MzEF"
/evidence-not_experimental
complement(18255..20096)
/rpt_family="Tigger1"
complement(18382..18510)
/note="GRAIL"
/evidence-not_experimental
complement(19602..19747)
/note="GRAIL"
/evidence-not_experimental
complement(19602..19747)
/note="GENSCAN"
complement(20169..20407)
/rpt_family="MIR"
complement(20575..20701)
/note="GRAIL"
/evidence-not_experimental
22106..22399
/rpt_family="AluX"
complement(22402..22449)
/note="(CATAT)n"
/rpt_family="Simple_repeat"
complement(22448..22503)
/note="(CATAT)n"
/rpt_family="Simple_repeat"
22994..23212
/rpt_family="LIMA3"
complement(23501..23948)
/rpt_family="LIME1"
23969..23989
/note="AT_rich"
/rpt_family="Low_complexity"
complement(24825..25242)
/rpt_family="LIMA6"
complement(26521..26547)
/note="AT_rich"
/rpt_family="Low_complexity"
26816..26850
/note="AT_rich"
/rpt_family="Low_complexity"
complement(27296..27855)
/rpt_family="LIMA2"
27952..28079
/rpt_family="LIMD_Send"
28098..28445
/rpt_family="LIMD_Send"
28471..28643
/rpt_family="LIM4_orf2"
28663..28846
/rpt_family="LIM4_orf2"
28857..29228
/rpt_family="LIM4_orf2"
complement(29236..29746)
/rpt_family="MER74A"
complement(29311..29369)
/note="GENSCAN"
29761..30627
/rpt_family="LIMD1"
29874..29921
/note="(TA)n"
/rpt_family="Simple_repeat"
30468..30584
/note="GRAIL"
/evidence-not_experimental
30642..30684
/note="AT_rich"
/rpt_family="Low_complexity"
30706..30755
/note="AT_rich"
/rpt_family="Low_complexity"
complement(30843..30927)
/note="GRAIL"
/evidence-not_experimental
```

```
repeat_region 30979..31064
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(30980..31031)
/note="(CA)n"
/rpt_family="Simple_repeat"
complement(31080..31211)
/note="GRAIL"
/evidence-not_experimental
complement(31086..31194)
/note="MzEF"
/evidence-not_experimental
complement(31086..31211)
```

Query Match 9.1%; Score 18; DB 40; Length 88481;
Best Local Similarity 100.0%; Pred.No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TTAGGGAATTTTGGTTG 74
|||||

Db 17960 TTAGGGAATTTTGGTTG 17943

Search completed: April 6, 2000, 21:16:15
Job time: 50360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 21:16:15 ; Search time 9481.22 Seconds
(without alignments)
-43.875 Million cell updates/sec

Title: US-09-090-672B-11
Perfect score: 137
Sequence: 1 CCACCGACCTGGCTGATGC.....TCGTTCAGTCCTGGTCTCT 137

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 821193 seqs, -1518192014 residues
Word size : 0
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pli:*
8: gb_pl2:*
9: gb_pri:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_vi:*
17: em_fun:*
18: em_hum1:*
19: em_hum2:*
20: em_in:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pli:*
27: em_ro:*
28: em_sts:*
29: em_sy:*
30: em_un:*
31: em_vi:*
32: gb_htg1:*
33: gb_htg2:*
34: gb_inl:*
35: gb_in2:*
36: em_ba1:*
37: em_ba2:*
38: em_hum3:*
39: em_hum4:*
40: gb_pr4:*
41: gb_htg3:*
42: gb_htg4:*
43: gb_htg5:*

44: gb_htg6:*
45: gb_htg7:*
46: em_htg1:*
47: em_htg2:*
48: em_htg3:*
49: em_hum5:*
50: gb_pl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	137	100.0	150228	40	AC003071	AC003071 Homo sapi
C 2	20	14.6	22287	45	AC017236	AC017236 Drosophil
C 3	20	14.6	78154	10	AC002466	AC002466 Human BAC
C 4	20	14.6	89743	12	AC003062	AC003062 Mouse Chr
C 5	20	14.6	149625	42	AC010025	AC010025 Drosophil
C 6	20	14.6	174063	41	AC010143	AC010143 Homo sapi
C 7	20	14.6	177591	33	AC006563	AC006563 Drosophil
C 8	20	14.6	180513	44	AC008020	AC008020 Mus muscu
C 9	20	14.6	189870	42	AC010104	AC010104 Homo sapi
C 10	19	13.9	7080	9	AB024745	AB024745 Homo sapi
C 11	19	13.9	50511	11	AC005214	AC005214 Homo sapi
C 12	19	13.9	86329	11	HS403M6	AL033392 Human DNA
C 13	19	13.9	135855	10	HS135E14	AJ010598 Homo sapi
C 14	19	13.9	135450	40	AC005951	AC005951 Homo sapi
C 15	19	13.9	157176	40	AC007092	AC007092 Homo sapi
C 16	19	13.9	160798	44	AC012536	AC012536 Homo sapi
C 17	19	13.9	172048	10	HS179N16	Z95152 Homo sapien
C 18	19	13.9	204027	42	AC007158	AC007158 Homo sapi
C 19	18	13.1	270	9	HUMVIPHM1	M14618 Human vasoa
C 20	18	13.1	474	13	G24450	G24450 human STS W
C 21	18	13.1	5190	11	HSU66615	U66615 Human SWI/S
C 22	18	13.1	14837	41	AC008821	AC008821 Homo sapi
C 23	18	13.1	23536	11	AC002120	AC002120 Homo sapi
C 24	18	13.1	32602	10	U73643	U73643 Human Chrom
C 25	18	13.1	35967	43	AC014889	AC014889 Drosophil
C 26	18	13.1	37027	11	HS506G2B	Z82976 Human DNA s
C 27	18	13.1	44294	10	AC000394	AC000394 Genomic s
C 28	18	13.1	48925	42	AC012318	AC012318 Homo sapi
C 29	18	13.1	55707	41	AC011108	AC011108 Homo sapi
C 30	18	13.1	81542	7	AB026647	AB026647 Arabidops
C 31	18	13.1	81874	11	HS931E15	AL023575 Human DNA
C 32	18	13.1	86765	11	AC004472	AC004472 Homo sapi
C 33	18	13.1	89328	11	HS398C22	Z93784 Homo sapien
C 34	18	13.1	89818	11	AC002126	AC002126 Homo sapi
C 35	18	13.1	93841	41	AC009012	AC009012 Homo sapi
C 36	18	13.1	95425	32	HSJ508G21	Z82213 Homo sapien
C 37	18	13.1	96099	41	AC011336	AC011336 Homo sapi
C 38	18	13.1	100091	44	AC013370	AC013370 Homo sapi
C 39	18	13.1	100997	11	HS262D12	Z99297 Homo sapien
C 40	18	13.1	109210	11	HSJ7663C7	AL049761 Human DNA
C 41	18	13.1	109891	10	HS112K5	Z85987 Human DNA s
C 42	18	13.1	110000	41	AC010679_2	Continuation (3 of
C 43	18	13.1	110000	43	AC015845_0	AC015845 Homo sapi
C 44	18	13.1	110000	43	AC015845_2	Continuation (3 of
C 45	18	13.1	121345	43	AC013673	AC013673 Homo sapi

ALIGNMENTS

RESULT 1
AC003071/c
LOCUS
DEFINITION Homo sapiens BAC clone BK085E05 from 22q12.1-qter, complete sequence.
AC003071
AC003071 150228 bp DNA PRI 18-MAR-1999

```

VERSION AC003071.1 GI:2588643
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150228)
AUTHORS Sulston, J.E. and Waterston, R.
JOURNAL toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE 2 (bases 1 to 150228)
AUTHORS Bradshaw, H., Graves, T., Sutterer, C. and Ozersky, P.
JOURNAL The sequence of Homo sapiens BAC clone BK085E05
MEDLINE Unpublished (1999)
REFERENCE 3 (bases 1 to 150228)
AUTHORS Waterston, R.
JOURNAL Direct Submission
TITLE Submitted (06-NOV-1997) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 150228)
AUTHORS Waterston, R.
JOURNAL Direct Submission
TITLE Submitted (18-MAR-1999) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the chromosome 22 mapping group
at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22/

SOURCE INFORMATION:
This clone is from the human BAC library described by U-J. Kim et
al., Genomics 34:213-8 (1996). This clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBAC108L
Selection: Chloramphenicol
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is DJ0394A18, 200 bp overlap; the
clone sequenced to the right is DJ43804, 200 bp overlap. Actual
start of this clone is at base position 1 of BK085E05.
Location/Qualifiers
1. .150228
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q12.1-qter"
/clone_lib="BKA"
/clone="BK085E05"
<19. .17848
/gene="WUGSC:H_BK085E05.1"

join(<19. .79.1065. .1176.4415. .4610.5403. .5588.8575. .8745.
9974. .10119.11327. .12931.14076. .14186.14656. .14744.
17677. .17848)
/gene="WUGSC:H_BK085E05.1"
/notes="match to AB002369 (NID:g2224682): H_BK085E05.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB83949.1"
/db_xref="GI:2588644"
/translation="ARSYAAVAANRAKGGGCEPEYYPNCEVFMGMANHISIRRSFQ
SLRLCTQMPDGNWLSAAESTKWLHLHLSALLVYHAVDQRPVLVHCSGDWD
RTPQIVALKLLDPYRTIEGQVLVEMEWLDFGHKADRCGCHGNSDDKNERCPVF
LQWLDCVHQLORQFPSCFENEAFYKLVQHTYSCLEFGLCNNAERKHHQERTC
SWSLLRAGNAFKNLISQSEAVLIPVCHVRNLMWLSAVILPCFSPITPVDDSCAP
YPAGTSPDDPLSADSLGKPTRVPGAEALSVAAGVAGGOMENILOEATKEESGVPEPAH
SSLAGCEGKEDPLLEKESRRKTPPEASGLHQDPELGDAAALSHLDSMWPLFSQGIS
RAGQIGISVLSSLIQVPRGEDSLEVPVEQFRIEIAEGREEAVLPVPDAKVCYGTQS
SCSLLPSQVFPETRGPNVSDTDLVEDKVSQSGQGHRSCLVNSGKDRLPQTMEP
SPSETSLVERQVGSVYHRTISUGTSLTSLRSPCALPLAECKEGLVNCAGPENRASE
QPGELSTLQMYPTPNGCHGANGAEGRSKLSRLSAMSCSSAHLHRNHHKWLHSHS
GRPSATSSPDQSRSHLDDGMSVYDTIQORLRQIESGHQOEVELKKQVQLKSLR
ESQYLTSLLHFNDFGDEVTSIPDSNLNDONCLSRCSTEIFSEASWEQVDKQDTEM
RWLPDLAAHCYACDSAFWLASRKHRCNCGNVFCSSCCNKQKVPVPSQQLFERSVCK
SCVSLHP7PSSSIDLELDKPIATSN"
2255. .2318
/rpt_family="MIR"
repeat_region
2968. .3065
/rpt_family="L2"
repeat_region
3427. .3709
/rpt_family="Alu"
repeat_region
3973. .4001
/rpt_family="(GAAA)n"
misc_feature
4415. .4611
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10292 (NID:g762248) yf36e10.r1"
misc_feature
5401. .5577
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10292 (NID:g762248) yf36e10.r1"
repeat_region
5855. .6544
/rpt_family="L2"
repeat_region
6554. .6853
/rpt_family="Alu"
repeat_region
6854. .7094
/rpt_family="L2"
repeat_region
7103. .7406
/rpt_family="Alu"
repeat_region
7418. .8020
/rpt_family="L2"
repeat_region
8109. .8221
/rpt_family="L2"
misc_feature
8575. .8745
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST H55106
complement(8768. .8869)
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10293
complement(8768. .8864)
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10293
complement(8831. .9183)
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10293
complement(8839. .9183)
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST R10293
9636. .10089
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST H75577
10302. .10410
/gene="WUGSC:H_BK085E05.1"
/notes="match to EST AA229246 (NID:g1851060) nc46g08.s1"
10507. .10803
repeat_region

```

FEATURES

source

gene

sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7> or send <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:

This clone is from a release of the human BAC library. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RG354L07; actual end is at 78154 of RG354L07. The orientation of this clone is unknown.

This clone contains STS's SWSS2071 (NID:g1113356) and SWSS2179 (NID:g1113411).

FEATURES

Source	Location/Qualifiers
	1..78154
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="RG354L07"
	/clone.lib="CITB-HS-A"
	/map="7q31"
repeat_region	complement(575..827)
repeat_region	/rpt_family="ALU"
repeat_region	1985..2017
repeat_region	/rpt_family="L1"
repeat_region	3425..3629
repeat_region	/rpt_family="MER"
repeat_region	3721..3940
repeat_region	/rpt_family="MER"
repeat_region	complement(4181..4594)
repeat_region	/rpt_family="MER"
repeat_region	complement(4625..4923)
repeat_region	/rpt_family="L1"
repeat_region	complement(4939..5482)
repeat_region	/rpt_family="L1"
repeat_region	5991..6283
repeat_region	/rpt_family="ALU"
repeat_region	complement(6957..7181)
repeat_region	/rpt_family="MER"
repeat_region	complement(7430..7453)
repeat_region	/rpt_family="L1"
repeat_region	7957..7992
repeat_region	/rpt_family="L1"
repeat_region	10650..10692
repeat_region	/rpt_family="L1"
repeat_region	12745..12817
repeat_region	/rpt_family="L1"
repeat_region	13582..13599
repeat_region	/rpt_family="L1"
repeat_region	14112..14318
repeat_region	/rpt_family="L1"
misc_feature	16949..17210
misc_feature	/note="match to EST N73830 (NID:g1231115) za56c01.s1"
repeat_region	complement(18244..18641)
repeat_region	/note="match to EST W00873 (NID:g1272993) za56c01.r1"
repeat_region	20048..20337
repeat_region	/rpt_family="ALU"

repeat_region	20338..20367	/rpt_family="L1"
repeat_region	21586..21600	/rpt_family="L1"
repeat_region	22318..22608	/rpt_family="ALU"
repeat_region	25037..25083	/rpt_family="L1"
repeat_region	complement(28356..28928)	/rpt_family="L1"
repeat_region	33433..33459	/rpt_family="L1"
repeat_region	34370..34460	/rpt_family="L1"
repeat_region	34718..34737	/rpt_family="L1"
repeat_region	complement(35485..35775)	/rpt_family="ALU"
repeat_region	36208..37412	/rpt_family="L1"
repeat_region	complement(40472..40559)	/rpt_family="L1"
repeat_region	40482..40507	/rpt_family="L1"
repeat_region	40560..40854	/rpt_family="ALU"
repeat_region	complement(40874..42045)	/rpt_family="L1"
repeat_region	complement(45725..45762)	/rpt_family="L1"
repeat_region	47435..47710	/rpt_family="ALU"
repeat_region	48328..48357	/rpt_family="L1"
repeat_region	complement(49283..49415)	/rpt_family="L1"
repeat_region	complement(54609..54901)	/rpt_family="ALU"
repeat_region	complement(57040..57069)	/rpt_family="L1"
repeat_region	59551..59571	/rpt_family="L1"
repeat_region	59576..59705	/rpt_family="ALU"
repeat_region	complement(60076..60338)	/rpt_family="ALU"
repeat_region	complement(60918..60945)	/rpt_family="L1"
repeat_region	complement(60954..61254)	/rpt_family="ALU"
repeat_region	complement(62605..62633)	/rpt_family="L1"
repeat_region	63387..63437	/rpt_family="L1"
repeat_region	complement(63476..63766)	/rpt_family="ALU"
repeat_region	67301..67331	/rpt_family="L1"
repeat_region	complement(68442..68462)	/rpt_family="L1"
repeat_region	complement(73916..74206)	/rpt_family="ALU"
repeat_region	complement(76596..76628)	/rpt_family="L1"
BASE COUNT	25369 a 13876 c 14122 g 24787 t	
ORIGIN		

Query Match 14.6%; Score 20; DB 10; Length 78154;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TCCAAGCTAAACAATTC 104

```
Db 75221 TCCAAGCTAAACAATTCTC 75240
|||||
AC003062 89743 bp DNA ROD 03-MAY-1999
Mouse Chromosome 16 Region Syntenic to DSCR BAC Clone b264n1,
complete sequence.
AC003062
AC003062.2 GI:4731672
HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 89743)
AUTHORS Chen,F. and Roe,B.A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89743)
AUTHORS Galli,N., Baldwin,S., Lund,J., Reeves,R., Gong,W., Chen,F.,
Roe,B.A., Emanuel,B.S., Nayak,S., Mickanin,C., Budarf,M.L. and
Buck,C.A.
TITLE A Region of Mouse Chromosome 16 is Syntenic to the DiGeorge,
Velo-Cardio-Facial Syndrome Minimal Critical Region
JOURNAL Unpublished
REMARK The genes were identified by comparing with human genomic and cDNA
sequences and RT-PCR of 12 day post conception mouse embryos total
RNA
REFERENCE 3 (bases 1 to 89743)
AUTHORS Roe,B.A. Dr.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1997) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 89743)
AUTHORS Roe,B.A. Dr.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 89743)
AUTHORS Roe,B.A. Dr.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1998) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
REFERENCE 6 (bases 1 to 89743)
AUTHORS Roe,B.A. Dr.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
REFERENCE 7 (bases 1 to 89743)
AUTHORS Roe,B.A. Dr.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-1999) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
COMMENT On May 3, 1999 this sequence version replaced gi:3845373.
FEATURES
source
1. .89743
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="b264n1"
BASE COUNT 24078 a 18138 c 20153 g 27374 t
ORIGIN
Query Match 14.6%; Score 20; DB 12; Length 89743;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 TCTTCTCTGAAGTCTTCCA 88
|||||
Db 61998 TCTTCTCTGAAGTCTTCCA 61999
```

RESULT 5

AC010025/c

LOCUS

DEFINITION

AC010025

AC010025.4

GI:6056162

HTG; HTGS_PHASE1.

SOURCE

ORGANISM

Drosophila melanogaster

Eukaryota; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 149625)

AUTHORS

Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,

Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,

Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,

David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,

Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,

Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,

Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,

Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,

Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,

Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,

Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,

Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,

Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,

Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,

Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P.,

Taylor,I., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,

Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,

Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.

Direct Submission

Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 149625)

AUTHORS

Worley,K.C.

TITLE

Direct Submission

JOURNAL

Submitted (11-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

COMMENT On Oct 16, 1999 this sequence version replaced gi:5916438.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 57 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 757: contig of 757 bp in length

* 758

* 2021: contig of 1264 bp in length

* 2022

* 2808: contig of 787 bp in length

* 2809

* 3570: contig of 762 bp in length

* 3571

* 4338: contig of 768 bp in length

* 5501: contig of 1163 bp in length

* 6340: contig of 839 bp in length

* 7057: contig of 717 bp in length

* 8022: contig of 965 bp in length

* 8023

* 8867: contig of 845 bp in length

* 8868

* 10294: contig of 1427 bp in length

* 10295

* 11903: contig of 1609 bp in length

* 11904

* 13083: contig of 1180 bp in length

* 13084

* 13947: contig of 864 bp in length

* 13948

* 15429: contig of 1482 bp in length

* 15430

* 17810: contig of 2381 bp in length

* 17811

* 19508: contig of 1699 bp in length

* 19510

* 21388: contig of 1879 bp in length

* 21389

* 23163: contig of 1775 bp in length

* 21389

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE JOURNAL

Submitted (12-FEB-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Aug 2, 1999 this sequence version replaced gi:5630021.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 689: contig of 689 bp in length
* 690 769: gap of unknown length
* 770 1401: contig of 632 bp in length
* 1402 1481: gap of unknown length
* 1482 2084: contig of 603 bp in length
* 2085 2164: gap of unknown length
* 2165 2738: contig of 574 bp in length
* 2739 2818: gap of unknown length
* 2819 3763: contig of 945 bp in length
* 3764 3843: gap of unknown length
* 3844 6770: contig of 2927 bp in length
* 6771 6850: gap of unknown length
* 6851 10859: contig of 4019 bp in length
* 10870 10949: gap of unknown length
* 10950 24667: contig of 13718 bp in length
* 24668 24747: gap of unknown length
* 24748 41201: contig of 16454 bp in length
* 41202 41281: gap of unknown length
* 41282 68640: contig of 27359 bp in length
* 68641 68720: gap of unknown length
* 68721 95333: contig of 26613 bp in length
* 95334 95413: gap of unknown length
* 95414 125111: contig of 29698 bp in length
* 125112 125191: gap of unknown length
* 125192 177301: contig of 52110 bp in length
* 177302 177381: gap of unknown length
* 177382 177591: contig of 210 bp in length.

FEATURES source

1. 177591
/organism="Drosophila melanogaster"
/strain="v: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="69F1-69F2"
/clone="BACR48A05 (D477) RPCI-98 48.A.5"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACE3.6")

BASE COUNT 52325 a 35740 c 35243 g 53234 t 1049 others
ORIGIN

Query Match 14.6% Score 20; DB 33; Length 177591;
Best Local Similarity 100.0%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Oy 20 CTTTCTATCTGACTTCTTT 39
|||||

Db 22042 CTTTCTATCTGACTTCTTT 22023
|||||

RESULT 8
AC008020

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC008020 180513 bp DNA HTG 23-NOV-1999
Mus musculus, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.
AC008020
AC008020.22 GI:64566790
HTG; HTGS-PHASE1.
house mouse.
Mus musculus

REFERENCE

1 (bases 1 to 180513)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS
TITLE
JOURNAL

Deschamps,S., Oomen,S., Draber,R., Becat,C. and Roe,B.A.
Unpublished
2 (bases 1 to 180513)

REFERENCE
AUTHORS
TITLE
JOURNAL

Deschamps,S., Oomen,S. and Roe,B.A.
Direct Submission
Submitted (10-JUL-1999) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT

On Nov 21, 1999 this sequence version replaced gi:6403511.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2144: contig of 2144 bp in length
* 2145 2211: gap of unknown length
* 2212 10395: contig of 8184 bp in length
* 10396 10462: gap of unknown length
* 10463 26456: contig of 15994 bp in length
* 26457 26523: gap of unknown length
* 26524 32813: contig of 6290 bp in length
* 32814 32880: gap of unknown length
* 32881 50735: contig of 17854 bp in length
* 50735 50802: gap of unknown length
* 50802 73104: contig of 22303 bp in length
* 73105 73171: gap of unknown length
* 73172 80446: contig of 7275 bp in length
* 80447 80512: gap of unknown length
* 80513 111028: contig of 30515 bp in length
* 111028 111094: gap of unknown length
* 111094 138342: contig of 27249 bp in length
* 138343 138408: gap of unknown length
* 138409 180513: contig of 42105 bp in length.

FEATURES source

1. 180513
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 51816 a 36189 c 37396 g 54499 t 613 others
ORIGIN

Query Match 14.6% Score 20; DB 44; Length 180513;
Best Local Similarity 100.0%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

Oy 69 TCTTCTCTGAAGTCTTCCA 88
|||||

Db 46125 TCTTCTCTGAAGTCTTCCA 46144
|||||

RESULT 9

AC010104

LOCUS
DEFINITION

AC010104 189870 bp DNA HTG 27-OCT-1999
Homo sapiens clone NH0540C18, *** SEQUENCING IN PROGRESS ***, 2
unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE

AC010104 GI:6136424
AC010104.2 GI:6136424
HTG; HTGS-PHASE1.
human.

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 189870)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189870)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Oct 27, 1999 this sequence version replaced gi:5870287.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 71164: contig of 71164 bp in length
* 71165 71182: gap of unknown length
* 71183 189870: contig of 118688 bp in length.
FEATURES             Location/Qualifiers
     source           1..189870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0540C18"
BASE COUNT  51911 a 40881 c 41550 g 55510 t 18 others
ORIGIN
Query Match      14.6%; Score 20; DB 42; Length 189870;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCAGCTGCTGATGC 20
      |||||
Db 23673 CCACCGCAGCTGCTGATGC 23692

RESULT 10
AB024745/c
LOCUS AB024745 7080 bp DNA PRI 09-JUN-1999
DEFINITION Homo sapiens gene for Fe65L2, complete cds.
ACCESSION AB024745
VERSION AB024745.1 GI:5019575
KEYWORDS Fe65L2.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (sites)
AUTHORS Tanahashi,H. and Tabira,T.
TITLE Genome structure and chromosomal mapping of the gene for Fe65L2
JOURNAL Interacting with Alzheimer's beta-amyloid precursor protein
MEDLINE Biochem. Biophys. Res. Commun. 258 (2), 385-389 (1999)
REFERENCE 99262108
AUTHORS Tanahashi,H.
TITLE 2 (bases 1 to 7080)
JOURNAL Direct Submission
JOURNAL Submitted (10-MAR-1999) to the DBJ/EMBL/GenBank databases. Hiroshi
Tanahashi, National Institute of Neuroscience, Division of
Demyelinating Disease and Aging; 4-1-1 Ogawahigashi, Kodaira, Tokyo
187-8502, Japan (E-mail:tanahashen@ncnp.go.jp,
Tel:81-42-341-2711(ex:5163), Fax:81-42-346-1747)
FEATURES             Location/Qualifiers
     source           1..7080
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
```

```
misc_feature 1018
/note="transcription start site"
misc_feature 1052
/note="transcription start site"
gene 1105..6764
/gene="Fe65L2"
join(1105..1153,1435..1598,1680..1756,2626..2886,
2850..2996,3122..3249,3500..3505,3648..3762,3900..3984,
4248..4331,4595..4710,4845..5036,6528..6764)
/gene="Fe65L2"
/function="interacting with Alzheimer's beta-amyloid
precursor protein"
/codon_start=1
/product="Fe65L2"
/db_xref="GI:5019576"
/translation="MLCKDYLAILVNCDDDLGWDHSLVEAGLPPGWRKIHDAACT
YWHVPSGSQWQRPRTWELGDAEDPGTGTGWLGRPPKGRSSLSLSDRSNSLSW
YGESYIQSMPEPGAKFAVRSLSGWVEFEDLAPGRKSIYVNNICIQQLAQTRSRQPP
DGWGECONMLMLKDKDMSLVNPLHSLHCQPLVHVRVWGSSGGRDRDFAVAS
DKDSMLKCHVRCQDVPKATIASALHGLCAQILSERVEVSGDASCCSPDPISPDLP
QVELLDVSAQAQAEALYNGTLPVTKAMGMDVLEAIGTLTARGDNRNAYPTMLSVS
DSLMTAHPIOAEASTEERPELWQCPVRLVTFIGVGRDPTFTGLIADLGROFQCAAFWC
OPHAGGLSEAVQAACMYQKCLVASAARGKANGAQAARLRLKRTSSMSDPSGGPLPL
PLLKGGVGAGATPRKGVFSFLDAFRKPSLLHMP"
polyA_signal 7058..7063
BASE COUNT  1464 a 1789 c 2017 g 1810 t
ORIGIN
Query Match      13.9%; Score 19; DB 9; Length 7080;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 GATCACCTCTGTCAGTCC 129
      |||||
Db 2763 GATCACCTCTGTCAGTCC 2745

RESULT 11
AC005214/c
LOCUS AC005214 50511 bp DNA PRI 01-JUL-1998
DEFINITION Homo sapiens chromosome 5, BAC clone 319C17 (LBNL H159), complete
sequence.
ACCESSION AC005214
VERSION AC005214.1 GI:3282166
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 50511)
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 50511)
AUTHORS Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 50511)
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
```



```

misc_feature      complement(27838..27936)
                  /note="GRAIL 2 excellent exon, frame 0"
prim_transcript  join(29553..29612,29893..30060,30138..30215,31580..31709,
32097..32224,32706..32798,33054..33177,33302..33498,
35102..35202)
                  /standard_name="Fe65L2"
                  /note="80%-91% identity rat Y13413 transcriptional
activator FE65"
repeat_region    complement(30471..30874)
                  /rpt_family="Alu"
misc_feature      31341..31456
                  /note="GRAIL 2 excellent exon, frame 1"
misc_feature      31582..31709
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      32108..32222
                  /note="GRAIL 2 excellent exon, frame 2"
misc_feature      32684..32791
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      33055..33206
                  /note="GRAIL 2 excellent exon, frame 2"
repeat_region    34605..34751
                  /rpt_family="MIR"

Query Match      13.9%; Score 19; DB 11; Length 50511;
Best Local Similarity 100.0%; Pred. No. 1.5; 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;

QY 111 GATCACCTCTGTCAGTCC 129
      |||||
Db 31223 GATCACCTCTGTCAGTCC 31205

RESULT 12
HS403M6/c
LOCUS
DEFINITION
      HS403M6 86329 bp DNA PRI 23-NOV-1999
      Human DNA sequence from clone 403M6 on chromosome 6q24.1-25.2.
      Contains two unconnected exons of the gene for Myasthenia Gravis
      autoantigen Gravin, and ESTs, STSS and GSSs, complete sequence.
ACCESSION
      AL033392
VERSION
      AL033392.5 GI:4826487
KEYWORDS
      HTG: Gravin; Myasthenia Gravis autoantigen.
SOURCE
      human.
ORGANISM
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
      1 (bases 1 to 86329)
      Williams, S.
      Direct Submission
      Submitted (20-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
      CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
      requests: clonerequest@sanger.ac.uk
      On May 13, 1999 this sequence version replaced gi:4678468.
      During sequence assembly data is compared from overlapping clones.
      Where differences are found these are annotated as variations.
      Together with a note of the overlapping clone name. Note that the
      variation annotation may not be found in the sequence submission
      corresponding to the overlapping clone, as we submit sequences with
      only a small overlap as described above.
      The following abbreviations are used to associate primary accession
      numbers given in the feature table with their source databases:
      Eni., EMBL; Swi., SWISSPROT; Tr., TREMBL
      This sequence is the entire insert of clone 403M6. This sequence
      has been finished according to sequence map criteria as follows. An
      attempt is made to resolve all sequencing problems, such as
      compressions and repeats, but not necessarily within known
      annotated human repeat sequence elements (e.g. Alu). Where the
      sequence is ambiguous, there is an annotation using the 'unsure'
      feature key.
      This sequence was generated from part of bacterial clone contigs of
      human chromosome 6, constructed by the Sanger Centre Chromosome 6
      Mapping Group. Further information can be found at
      http://www.sanger.ac.uk/HGP/Chr6
      403M6 is from the library RC13 constructed at the Roswell Park

```

```

Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCIPAC2.
FEATURES
Source
1..86329
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q24.1-25.2"
/clone="RP3-403M6"
/clone_lib="RPCI-3"
201..494
repeat_region
/note="AluSp repeat: matches 1..299 of consensus"
550..832
repeat_region
/note="AluJb repeat: matches 1..300 of consensus"
1672..1982
repeat_region
/note="AluSx repeat: matches 1..312 of consensus"
2091..2330
repeat_region
/note="L1PA16 repeat: matches 5943..6157 of consensus"
2331..2630
repeat_region
/note="AluSg repeat: matches 1..300 of consensus"
2631..2763
repeat_region
/note="L1PA16 repeat: matches 5804..5943 of consensus"
3673..3960
repeat_region
/note="AluSx repeat: matches 1..305 of consensus"
3961..3982
repeat_region
/note="11 copies 2 mer tg 100% conserved"
4022..4061
repeat_region
/note="20 copies 2 mer at 80% conserved"
4097..4379
repeat_region
/note="AluSx repeat: matches 1..282 of consensus"
4434..4545
repeat_region
/note="56 copies 2 mer at 83% conserved"
4546..4579
repeat_region
/note="17 copies 2 mer ac 100% conserved"
4605..4910
repeat_region
/note="AluJb repeat: matches 1..307 of consensus"
5749..5994
repeat_region
/note="AluSx repeat: matches 5..251 of consensus"
6140..6432
repeat_region
/note="AluSx repeat: matches 1..305 of consensus"
6443..6748
repeat_region
/note="AluSx repeat: matches 1..297 of consensus"
6780..6877
repeat_region
/note="AluSp/q repeat: matches 215..312 of consensus"
7006..7288
repeat_region
/note="AluSp repeat: matches 1..291 of consensus"
7289..7765
misc_feature
/note="match: EST T71187"
7870..8149
misc_feature
/note="match: GSS B89849"
8079..8366
repeat_region
/note="AluJo repeat: matches 3..294 of consensus"
8607..8912
repeat_region
/note="AluSx repeat: matches 1..313 of consensus"
8913..8994
repeat_region
/note="L2 repeat: matches 1619..1698 of consensus"
complement(8992..9405)
misc_feature
/note="match: STS G27817"
complement(8995..9731)
misc_feature
/note="match: ESTs T72883 N58206"
9427..9722
repeat_region
/note="AluSx repeat: matches 5..300 of consensus"
9904..10199
repeat_region
/note="AluSx repeat: matches 1..297 of consensus"
10266..10343
repeat_region
/note="L1MB7 repeat: matches 6099..6173 of consensus"
10456..10841
misc_feature
/note="match: GSS AQ134006"
11936..12074
repeat_region
/note="AluY repeat: matches 167..307 of consensus"
13166..13471
repeat_region
/note="AluSp repeat: matches 2..303 of consensus"
13662..13977
repeat_region

```

```
/note="AluSg repeat: matches 1. .307 of consensus"
13978. .14013
repeat_region /note="AluJb repeat: matches 1. .312 of consensus"
repeat_region 32627. .33244
/note="L1PA7 repeat: matches 5520. .6138 of consensus"
33267. .33567
repeat_region /note="AluSg1 repeat: matches 1. .301 of consensus"
15090. .15391 complement(34525. .35194)
/note="AluSx repeat: matches 1. .303 of consensus" /note="match: GSS AQ268889"
15392. .15828 complement(34670. .35222)
/note="L1MB8 repeat: matches 5386. .5840 of consensus" /note="match: GSS AQ353171"
15461. .15904 /note="match: EST M62284"
16184. .16358
repeat_region /note="L1MC4 repeat: matches 7705. .7880 of consensus"
16360. .16659
repeat_region /note="AluSx repeat: matches 1. .298 of consensus"
16689. .16993
repeat_region /note="AluSx repeat: matches 1. .304 of consensus"
17093. .17402
repeat_region /note="AluJb repeat: matches 1. .306 of consensus"
/note="AluJb repeat: matches 1. .306 of consensus"
complement(17391. .17672)
/note="match: STS G08536 G09147"
17528. .17569
repeat_region /note="21 copies 2 mer ga 76% conserved"
17572. .17865
repeat_region /note="AluSx repeat: matches 1. .282 of consensus"
17866. .17957
repeat_region /note="L1MC4 repeat: matches 7413. .7505 of consensus"
17958. .18240
repeat_region /note="AluJb repeat: matches 1. .298 of consensus"
18241. .18284
repeat_region /note="L1MC4 repeat: matches 7370. .7413 of consensus"
18786. .19083
repeat_region /note="AluSg repeat: matches 5. .301 of consensus"
19108. .19412
repeat_region /note="AluJb repeat: matches 1. .298 of consensus"
19857. .20147
repeat_region /note="AluJb repeat: matches 3. .292 of consensus"
22576. .22624
repeat_region /note="L2 repeat: matches 2701. .2749 of consensus"
22979. .23014
repeat_region /note="18 copies 2 mer tg 94% conserved"
23238. .23549
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
23708. .23999
repeat_region /note="AluJ repeat: matches 2. .293 of consensus"
24736. .24898
repeat_region /note="FRAM repeat: matches 5. .161 of consensus"
25209. .25316
repeat_region /note="L1ME repeat: matches 5683. .5790 of consensus"
25549. .25646
repeat_region /note="U6 repeat: matches 1. .107 of consensus"
25669. .25839
repeat_region /note="FRAM repeat: matches 0. .175 of consensus"
26724. .27021
repeat_region /note="AluJb repeat: matches 11. .305 of consensus"
27122. .27443
repeat_region /note="AluSx repeat: matches 1. .296 of consensus"
27629. .27925
misc_feature /note="match: GSS B75668"
27677. .27683
misc_feature /note="IS150 excised."
28611. .28677
repeat_region /note="MIR repeat: matches 97. .166 of consensus"
29085. .29243
repeat_region /note="L1ME3A repeat: matches 5947. .6118 of consensus"
29378. .29689
repeat_region /note="AluJb repeat: matches 1. .307 of consensus"
30419. .30452
repeat_region /note="17 copies 2 mer tg 100% conserved"
30914. .31194
repeat_region /note="AluSg repeat: matches 18. .298 of consensus"
31963. .32272
repeat_region /note="AluJ repeat: matches 1. .309 of consensus"

repeat_region 32275. .32594
/note="AluJb repeat: matches 1. .312 of consensus"
32627. .33244
/note="L1PA7 repeat: matches 5520. .6138 of consensus"
33267. .33567
/note="AluSg1 repeat: matches 1. .301 of consensus"
complement(34525. .35194)
/note="match: GSS AQ268889"
35217. .35639
/note="match: GSS AQ353171"
/note="match: GSS AQ200990 AQ505480"
35640. .35943
repeat_region /note="AluSx repeat: matches 1. .300 of consensus"
36195. .36288

Query Match 13.9% Score 19: DB 11: Length 86329;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGATG 19
|||||
Db 41543 CCACCGCACCTGGCTGATG 41525

RESULT 13
HS135E14
LOCUS HS135E14 135855 bp DNA PRI 03-NOV-1999
DEFINITION Homo sapiens chromosome 21 PAC RPCIP704E1413Q02, complete sequence.
ACCESSION AJ010598
VERSION AJ010598.1 GI:3559851
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135855)
AUTHORS Brandt,P., Dose,S., Grimm,M., Groet,J., Hornischer,K.,
Loehner,T.H., Nizetic,D., Scharfe,M., Schoen,O., Yaspo,M.L. and
Bioecker,H.
Direct Submission
Submitted (30-AUG-1998) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
bioecker@gbf.de
Sequence overlaps with Acc.Nr. AJ010597
Collaborators:
Center for applied Molecular Biology
School of Pharmacy, University of London
29-39 Brunswick Square, London, WC1N 1AX, UK
e.mail: j.groet@hemb.uisop.ac.uk
and
GBF, Dept. of Genome Analysis,
Mascheroder Weg 1, D-38124 Braunschweig, Germany,
and
Max-planck-Institut fuer Molekulare Genetik
Innestrasse 73, D-14195 Berlin-Dahlem, Germany
All annotations in this database entry are developed by
computational tools. It is therefore not explicitly noted in the
feature lines that evidence is not experimental.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bioecker).
+ Programs used by 'AnnoMitter':
+++++
> GeneFinder (Green), Vers. 084
> Organism: human
> GenScan (Burge & Karlin), Vers. 1.0
> Used matrix: vertebrate; Minimum score: 0
> Grail (Xu et al.), Vers. 1.3
> Organism: human
> Mzef (Zhang)
```

```

: Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
& Skolnick)
: Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
BLASTN 2.0.9 (Altschul et al.)
: Database(s): * Repbase (human), released 22-DEC-1995
* RepBase (primate), released 22-DEC-1995
: Minimum score: 60;
Minimum identity: 70 %;
> 'ESTs': BLASTN 2.0.9 (Altschul et al.)
: Database(s): * embl (EST), Vers. 60 (16-SEP-1999)
* emblnew (EST), Vers. 60+ (20-OCT-1999)
: Using unmasked
sequence
: Minimum score: 60; Minimum identity: 70 %;
> 'GSSs': BLASTN 2.0.9 (Altschul et al.)
: Database(s): * embl (GSS), Vers. 60 (16-SEP-1999)
* emblnew (GSS), Vers. 60+ (20-OCT-1999)
: Using unmasked
sequence
: Minimum score: 60; Minimum identity: 70 %;
> 'Tandem Repeats': GDE 2.2 option 'candem'
: Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
: Treat N's as mismatches? YES; Allow uniform consensus? NO >
'Inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
Islands': GDE 2.2 option 'cpg'
: CpG island region size 100 bp;
: Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
e-PCR (Schuler)
: Margin: 50; Number of mismatches allowed: 0; Word size: 7
: STS database: 'dbSTS markers'
> 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
Location/Qualifiers
1..135855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="RPCIP704E14135Q2"
/map="21q11.2"
misc_feature
1..12050
/note="overlap bases 93949..105999 in AJ010597"
satellite
72..83
/note="AGC repeat"
misc_feature
complement(415..554)
/note="match: EST AA482201"
exon
417..554
/note="GENSCAN prediction, score = 4.94; GRAIL, score =
100%, comment = excellent; MZEF prediction, score = 0.666"
misc_difference
1736..1737
/note="insertion in AJ010597, pos. 95684..95686"
/misc="ttc"
misc_feature
complement((2237..2345)..(2300..2388))
/note="match: GSSs AQ095980 AQ171887 AQ387155 AQ139774
AQ356842 AQ355302 B76031 AQ518812 AQ663333 B80647 AQ729504
B53194 B57183 AQ505725 AQ035195 AQ822466 AQ307994 AQ081956
AQ003991 AQ761322 B84086 AQ300632 B95497 AQ744175 AQ100967
AQ231861 B55560 AQ009809 AQ406989 AQ000416 AQ263687
AQ658752 AQ552102 AQ343163 AQ705263 AQ629490 AQ793378
AQ733071 AQ458474 AQ374766 AQ041660 AQ392400 B90870
AQ368709 AQ326414 AQ416338 AQ518037 AQ122336 AQ534392
AQ353324 B46776 AQ631289 AQ587966 AQ356012 AQ830605 B86437
AQ103375 AQ336818 AQ479135 AQ475599 AQ379051 AQ183111
AQ228770 B42652 AQ371549 AQ456929 AQ213075 AQ486054"
/note="match: ESTs AW019895 AA810343 AL047793 AT027504
H47919 AA376278 AA335765 AA376189 W42477 AA053573 AA375658
AL048534 W93437 AA371935 AA490123 AA402704 T55145 AA486656
AA773990 AA283630 H67233 T59499 R50612 AA515631 AA523086
AA302982 AA480492 AA74199 T62882 AA744303 AA015725 T76991
AA558696 AA229904 AA483754 A1816100 A1355803 AA679952
AA745348 R19969 AA290966 AL046262 AA715348 AA234804
AA344085 F03310 AL120083"
misc_feature
((2252..2345)..(2304..2386))
/note="match: ESTs AW051819 H72628 AI608668 AI032875
AA181987 AA486559 AA244397 AI669239 AL043105 H49262

```

```

AI023401 AA761008 AA505142 AA829847 AA021034 AA310191
H09123 AA559030 AI871724 T72674 T98886 H81552 AA923712
H10668 AI076357 AT272192 AT368252 H61034 AI051619 H79348
AI473949 AA878105 AI832910 AA483936 AI083610 AA745638
AA10556 AL035988 AA385780 AI168100 AA877973 AI324602
AA878105 AA225230 AI453239 N41859 AA582842 AI953769
AI25049 R36445 AA180344 AA569591 AA631359 AI215714
AA643211 AI978999 AL046021 AA229905 AI168101 AA682635
AA716522 AA019068 AI81484 AL038528 N57913 AA180487
AA52596 AA315566 AL045829 AA100431 AA226153 AA775332
AA809926 AA487071 AI683635 AA469242 C75350 T93092 AA046440
AA642464 AA708753 AA534054 T17218 H61906 AI368836 AA909694
H90784 AA015590 AA720979 H84764 R02522 AA669369 H74442
AA614144 AA732435 AA604995 AI590789 AA59789 AA59613 AA491954
AA526639 AA113159 AW082490 AW006930 AW006919 AW007621
AW078909 AI992193 AW007989 AL045954 AA643964"
(2252..2341)..(2311..2397)
/note="match: GSSs AG020681 AQ422183 AQ395382 AQ543764
B92811 AQ103217 AQ525868 AQ032940 AQ528253 AQ662555
AQ136985 AQ190729 AQ472993 AQ267937 AQ473182 AQ551390
AQ527731 AQ042077 AQ484780 AQ550517 AQ167022 B87081
AQ269024 AQ630192 AQ007744 AQ176782 AQ198723 AQ701150
AQ175828 AQ506370 AQ350441 AQ143462 AQ533970 AQ385987
AQ208962 AQ347476 AQ745990 AQ556682 AQ529563 AQ490365
B68221 AQ530224 B05868 AQ490786 AQ054307 AQ507905 B60385
AQ681971 AQ489261 AQ343157 AQ389527 AQ310151 AQ737169
AQ549438 AQ829593 AQ231224 AQ478614 AQ375510 AQ227632
AQ134004 AQ429116 AQ207026 AQ780082 AQ719813 AQ002146
AQ115286 AQ570005"
2286..2345
/note="Genefinder prediction"
2388..2402
/note="ATTTT repeat"
complement(2399..2512)
/note="86% identity: matches 11..124 of consensus"
/rpt_family="AluJo"
2782..2847
/note="match: EST W91939"
complement(2798..2965)
/note="Genefinder prediction"
complement(2842..2965)
/note="GRAIL, score = 79%, comment = excellent shadow"
2967..2979
/note="TGTC repeat"
3032..3073
/note="match: GSS AQ535294"
3034..3085
/note="match: EST AI160319"
complement(3043..3076)
/note="match: EST AA642169"
3514..3581
/note="XPOUND prediction, score = 0.528"
3707..3732
/note="IR1, 84% complementary to IR1' (3750..3775)"
/rpt_type="INVERTED"
3750..3775
/note="IR1", 84% complementary to IR1 (3707..3732)"
/rpt_type="INVERTED"
3842..3868
/note="TG repeat"
3842..3867
/note="homology = 100.0%, counts = 13"
/rpt_family="tg repeat"
/rpt_type="TANDEM"
complement(4101..4141)
/note="match: EST AA482201"
4104..4180
/note="GENSCAN prediction, score = 3.89; GRAIL, score =
100%, comment = excellent; MZEF prediction, score = 0.996"
4138..4180
/note="XPOUND prediction, score = 0.939"
4258..4604
/note="match: GSS AQ728019"

```

```

misc_feature      4370..4708
/note="match: EST H69816"
(4786..4816)..(4822..4885)
misc_feature      /note="match: GSSs AQ542359 AQ544443 AQ426467 AQ126259
AQ20063 AQ546084 AQ205053 AQ322830 AQ195430 AQ337507
AQ221567 AQ065988 AQ594678 AQ036247 AQ542440 B34620"
(4786..4813)..(4836..4885)
misc_feature      /note="match: ESTs A1243353 A1357315 AA325371 A1338348
(4786..4813)..(4836..4885)

Query Match      13.9%; Score 19; DB 10; Length 135855;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 AGCTTCCAGCTAAACA 98
|||||
Db 16163 AGCTTCCAGCTAAACA 16181

RESULT 14
AC005951/c
LOCUS
DEFINITION Homo sapiens chromosome 17, clone hRPK.372_K_20, complete sequence.
ACCESSION AC005951
VERSION AC005951.1 GI:3992081
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155450)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 17, clone hRPK.372_K_20
Unpublished
2 (bases 1 to 155450)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Allen.N., Anderson.M.,
Baker.J., Baldwin.J., Barna.N., Beckerly.R., Benn.J., Boutwell.C.,
Brown.A., Castile.A., Cerny.J., Colangelo.M., Collins.S.,
Collamore.A., Cooke.P., Corliss.D., Depayre.E., Devon.K., Dewar.K.,
Donelan.L., Ferreira.P., FitzHugh.W., Forrest.C., Funke.R.,
Gage.D., Gardyna.S., Geraghty.K., Grant.G., Hagos.B., Heaford.A.,
Herena.L., Horton.L., Howland.J.C., Jacotot.L., Jones.C., Kann.L.,
Karatas.A., Lehoczyk.J., Macdonald.P., Marquis.N., McEwan.P.,
McGurk.A., McKernan.K., Meidrim.J., Molla.M., Morris.W., Morrow.J.,
Mychaleckyj.J., Nahf.R., Naylor.J., Niloff.M., O'Connor.T.,
O'Donnell.P., Pavlin.B., Peterson.K., Riley.R., Roberts.D., Roy.A.,
Severy.P., Stange-Thomann.N., Stilwell.J., Stojanovic.N., Stone.C.,
Subramanian.A., Tesfaye.S., Tichovolsky.N., Torruella-Miller.I.,
Vassiliev.H., Vo.A., Wagner.A., Wheeler.J., Wu.Y., Wyman.D.,
Ye.W.J., Zhao.J. and Zody.M.
Direct Submission
Submitted (07-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 155450)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Allen.N., Anderson.M.,
Baker.J., Baldwin.J., Barna.N., Beckerly.R., Benn.J., Boutwell.C.,
Brown.A., Castile.A., Cerny.J., Colangelo.M., Collins.S.,
Collamore.A., Cooke.P., Corliss.D., Depayre.E., Devon.K., Dewar.K.,
Donelan.L., Ferreira.P., FitzHugh.W., Forrest.C., Funke.R.,
Gage.D., Gardyna.S., Geraghty.K., Grant.G., Hagos.B., Heaford.A.,
Herena.L., Horton.L., Howland.J.C., Jacotot.L., Jones.C., Kann.L.,
Karatas.A., Lehoczyk.J., Macdonald.P., Marquis.N., McEwan.P.,
McGurk.A., McKernan.K., Meidrim.J., Molla.M., Morris.W., Morrow.J.,
Mychaleckyj.J., Nahf.R., Naylor.J., Niloff.M., O'Connor.T.,
O'Donnell.P., Pavlin.B., Peterson.K., Riley.R., Roberts.D., Roy.A.,
Severy.P., Stange-Thomann.N., Stilwell.J., Stojanovic.N., Stone.C.,
Subramanian.A., Tesfaye.S., Tichovolsky.N., Torruella-Miller.I.,
Vassiliev.H., Vo.A., Wagner.A., Wheeler.J., Wu.Y., Wyman.D.,
Ye.W.J., Zhao.J. and Zody.M.
Direct Submission
Submitted (18-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 1998 this sequence version replaced gi:3868745.
All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
Location/Qualifiers
source 1..155450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.372_K_20"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
repeat_region 1968..2273
/rpt_family="AluY"
repeat_region complement(3334..3431)
/rpt_family="MIR"
repeat_region 3490..3702
/rpt_family="MER58A"
repeat_region complement(3712..3812)
/rpt_family="L2"
repeat_region complement(3841..4131)
/rpt_family="MLT1D"
repeat_region complement(4198..4238)
/rpt_family="(TAAA)n"
repeat_region complement(4239..4522)
/rpt_family="AluJo"
repeat_region complement(4523..4612)
/rpt_family="L2"
repeat_region 6026..6172
/rpt_family="MIR"
repeat_region complement(6989..7017)
/rpt_family="(CA)n"
repeat_region complement(7219..8061)
/rpt_family="L2"
repeat_region 10963..12285
/rpt_family="L1PA4"
repeat_region 12295..12334
/rpt_family="(TAAA)n"
repeat_region 12594..12698
/rpt_family="MLT1J"
repeat_region complement(13005..13180)
/rpt_family="MIR"
repeat_region complement(15210..15265)
/rpt_family="L2"
repeat_region complement(15552..17481)
/rpt_family="L1PA2"
repeat_region 17479..18391
/rpt_family="L1PA2"
repeat_region complement(19533..19555)
/rpt_family="(CAAA)n"
repeat_region complement(19556..19838)
/rpt_family="AluSg"
repeat_region 20101..20121
/rpt_family="AT-rich"
repeat_region 20135..20462
/rpt_family="MLT1A1"
repeat_region 21131..22493
/rpt_family="L1MB8"
repeat_region 23468..23817
/rpt_family="L1MB8"
repeat_region 24700..25033
/rpt_family="L1M4"
repeat_region 25053..25172
/note="Single-stranded terminator coverage."
repeat_region 25183..25677
/rpt_family="L1M4"
repeat_region 25748..25856
/rpt_family="MLT1J"
repeat_region 25958..26130
/rpt_family="MLT1I"
repeat_region complement(26155..27023)
/rpt_family="L1P"
repeat_region 27025..27424
/rpt_family="L1PA13"
repeat_region 28022..28136

```

```

repeat_region /rpt_family="L1ME2"
28137..28444
repeat_region /rpt_family="AluY"
28445..28692
repeat_region /rpt_family="L1ME2"
28717..28761
repeat_region /rpt_family="AT_rich"
28876..29000
repeat_region /rpt_family="(TA)n"
complement(29647..29768)
repeat_region /rpt_family="L2"
30106..31622
repeat_region /rpt_family="L1PA2"
complement(31743..31945)
repeat_region /rpt_family="MLT1D"
complement(32064..32228)
repeat_region /rpt_family="AluJo/FRAM"
complement(32229..32359)
repeat_region /rpt_family="AluSc"
complement(32361..32902)
repeat_region /rpt_family="LTR12"
complement(32960..33008)
repeat_region /rpt_family="Alu"
complement(33159..33195)
repeat_region /rpt_family="(TAA)n"
complement(33196..33476)
repeat_region /rpt_family="AluSc"
34002..34353
repeat_region /rpt_family="MLT1A1"
complement(34355..34772)
repeat_region /rpt_family="NSTA"
complement(34874..34912)
repeat_region /rpt_family="(CAAAA)n"
complement(34913..35185)
repeat_region /rpt_family="AluSg"
35582..35650
repeat_region /rpt_family="AT_rich"
complement(36881..37168)
repeat_region /rpt_family="AluJo"
complement(37211..37558)
repeat_region /rpt_family="L1MB6"
38116..38245
repeat_region /rpt_family="L2"
38406..38446
repeat_region /rpt_family="AT_rich"
38474..40794
repeat_region /rpt_family="L1PA10"
41093..41115
repeat_region /rpt_family="AT_rich"
complement(41286..41333)
repeat_region /rpt_family="purine-rich"
41353..41386
repeat_region /rpt_family="(CA)n"
41781..41882
repeat_region /rpt_family="(TA)n"
42039..42080
repeat_region /rpt_family="AT_rich"
42439..42733
repeat_region /rpt_family="AluSx"
complement(43412..43600)
repeat_region /rpt_family="LTR41"
43822..43857
repeat_region /rpt_family="AT_rich"
43900..44072
repeat_region /rpt_family="L1MD1"
44151..44595
repeat_region /rpt_family="L1MD1"
44772..44816
repeat_region /rpt_family="AT_rich"
45430..45692
repeat_region /rpt_family="L1"
45693..46518
repeat_region /rpt_family="L1MA6"

```

```

repeat_region complement(46617..46812)

Query Match 13.9%; Score 19; DB 40; Length 155450;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGATG 19
    |||||
Db 52813 CCACCGCACCTGGCTGATG 52795

RESULT 15
AC007092
LOCUS AC007092 157176 bp DNA PRI 23-OCT-1999
DEFINITION Homo sapiens BAC clone NH0090D01 from 2, complete sequence.
AC007092
AC007092.4 GI:6102673
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 157176)
Toward a complete human genome sequence
Sulston, J.E. and Waterston, R.
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE
2 (bases 1 to 157176)
McPherson, C., Strommatt, C. and Davidson, T.
The sequence of Homo sapiens BAC clone NH0090D01
Unpublished
3 (bases 1 to 157176)
Waterston, R.H.
Direct Submission
Submitted (16-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 157176)
Waterston, R.H.
Direct Submission
Submitted (29-JUL-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 157176)
Waterston, R.
Direct Submission
Submitted (23-OCT-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 23, 1999 this sequence version replaced gi:5649381.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapienswatson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis

```


MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frenken,E., Tatenos,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACes3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is NH0179A20; the clone sequenced to the right is NH0260K08, 200 bp overlap. Actual end of this clone is at base position 15376 of NH0260K08.

FEATURES

source

1. 157176
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="NH0090D01"
/clone_lib="RPCI-11"
1692..2043
/rpt_family="MaLR"
2738..2780
/rpt_family="MER1_type"
2784..2965
/rpt_family="MER1_type"
2971..3306
/rpt_family="MER2_type"
4916..5060
/rpt_family="MER1_type"
5349..5586
/rpt_family="(TAGA)n"
5615..6801
/rpt_family="L1"
6812..6880
/rpt_family="MIR"
6881..7056
/rpt_family="MaLR"
7057..7343
/rpt_family="Alu"
7346..7367
/rpt_family="(TTAA)n"
7388..7596
/rpt_family="MaLR"
7597..7738
/rpt_family="MIR"
7769..7803
/rpt_family="(CTA)n"
8098..8126
/rpt_family="AT-rich"
8137..8228
/rpt_family="MIR"
10364..10684
/rpt_family="Alu"
11309..11456
/rpt_family="L2"
12538..12558
/rpt_family="AT-rich"
12564..12620
/rpt_family="MER1_type"
12639..12981
/rpt_family="MER1_type"
14025..14330
/rpt_family="Alu"
16136..16267
/rpt_family="MIR"
16316..16404
/rpt_family="MaLR"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

Query Match 13.9%; Score 19; DB 40; Length 157176;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCACCGCACCTGGCTGATG 19

16943..17093
/rpt_family="MER103"
17233..17610
/rpt_family="MaLR"
17611..19152
/rpt_family="MaLR"
19153..19528
/rpt_family="MaLR"
19575..19655
/rpt_family="MIR"
19867..20055
/rpt_family="MaLR"
20177..20207
/rpt_family="AT-rich"
20328..20464
/rpt_family="MIR"
21588..21816
/rpt_family="MIR"
22905..22954
/rpt_family="(CA)n"
24488..24517
/rpt_family="(TTTA)n"
24527..24732
/rpt_family="Alu"
25441..25468
/rpt_family="AT-rich"
26918..27105
/rpt_family="MER1_type"
28798..28976
/rpt_family="MIR"
28988..29865
/rpt_family="L1"
29916..30050
/rpt_family="MER1_type"
31276..31397
/rpt_family="MER1_type"
31548..31648
/rpt_family="MER1_type"
32045..32282
/rpt_family="MER1_type"
32289..32398
/rpt_family="MIR"
32505..32590
/rpt_family="(TC)n"
33437..33467
/rpt_family="(GGGTG)n"
36487..36538
/rpt_family="(CA)n"
38116..38143
/rpt_family="(TC)n"
38143..38179
/rpt_family="(TG)n"
38629..38988
/rpt_family="MaLR"
39148..39426
/rpt_family="Alu"
39478..39699
/rpt_family="MER1_type"
39722..39891
/rpt_family="MaLR"
40336..40455
/rpt_family="MER94"
40564..40725
/rpt_family="MaLR"
40836..40967
/rpt_family="MaLR"
41585..41624

sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The true left end of clone 179N16 is at 1 in this sequence. The true right end of clone 524E15 is at 690.
 The true left end of clone 179N16 is at 172048.
 179N16 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.
 For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES

```

Source
Location/Qualifiers
1..172048
/dborgasm="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p21.1-21.33"
/clone="RPI-179N16"
/clone_lib="RPC1-1"
/complement(1..1077)
/note="match: STS ALU20972"
4..457
/note="L1 repeat: matches 5082..4624 of consensus"
493..734
/note="AluJb repeat: matches 302..62 of consensus;
incomplete repeat"
1714..1934
/note="L1MB8 repeat: matches 324..574 of consensus"
1956..2257
/note="AluJb repeat: matches 1..302 of consensus"
2258..2543
/note="L1MB6 repeat: matches 566..874 of consensus"
2548..2848
/note="AluX repeat: matches 301..1 of consensus"
2879..3160
/note="AluJb repeat: matches 277..2 of consensus;
incomplete repeat"
3168..3369
/note="MER44C repeat: matches 258..463 of consensus"
4208..4497
/note="AluJb repeat: matches 289..1 of consensus"
5473..5894
/note="MLT1C repeat: matches 39..465 of consensus"
5986..6284
/note="AluJb repeat: matches 1..296 of consensus"
6361..6420
/note="AluX repeat: matches 301..242 of consensus;
incomplete repeat"
6612..6914
/note="AluJb repeat: matches 2..302 of consensus"
6951..7084
/note="FLAM_C repeat: matches 1..133 of consensus"
7621..7941
/note="L1 repeat: matches 5068..5387 of consensus"
7809..8170
/note="L1MC2 repeat: matches 15..384 of consensus"
8255..8378
/note="MER2 repeat: matches 13..136 of consensus"
8396..8603
/note="AluJb repeat: matches 291..82 of consensus;
incomplete repeat"
8605..8691
/note="MER2 repeat: matches 125..209 of consensus"
8826..8890
/note="MER2 repeat: matches 279..344 of consensus"
9518..9894
/note="THE1C repeat: matches 3..371 of consensus"
9964..9989
/note="13 copies of 2 mer 96 % conserved"
10970..11269
/note="AluSg repeat: matches 300..1 of consensus"
12207..12333
/note="MIR repeat: matches 106..248 of consensus"
12735..13031
/note="AluX repeat: matches 1..298 of consensus"
13042..13256
/note="AluSg repeat: matches 2..217 of consensus;
incomplete repeat"
13494..13581
/note="AluX repeat: matches 205..288 of consensus;
incomplete repeat"
13595..13953
/note="THE1B repeat: matches 364..1 of consensus"
13972..14648
/note="L1MC3 repeat: matches 1676..2345 of consensus"
14649..14948
/note="AluX repeat: matches 1..301 of consensus"
14949..15077
/note="L1MC3 repeat: matches 2336..2470 of consensus"
15102..15231
/note="FLAM_A repeat: matches 133..4 of consensus"
15244..15289
/note="23 copies of 2 mer 100 % conserved"
15674..15853
/note="match: 263849 CpG Island clone 91c11"
16341..16702
/note="MLT2FA repeat: matches 33..394 of consensus"
16823..16969
/note="MER11A repeat: matches 4..151 of consensus"
16968..17594
/note="MER11A repeat: matches 60..738 of consensus"
17278..17886
/note="MER11B repeat: matches 2..627 of consensus"
17977..18027
/note="MLT2FA repeat: matches 400..450 of consensus"
19217..19333
/note="MIR repeat: matches 262..117 of consensus"
19471..19693
/note="AluJb repeat: matches 85..301 of consensus;
incomplete repeat"
21094..21432
/note="MER1B repeat: matches 1..337 of consensus"
21701..22077
/note="match: ESTs R71395 R69756 AA587582 AA629092 R50348"
21706..21985
/note="match: STS G23312"
22054..22225
/note="AluJb repeat: matches 131..301 of consensus;
incomplete repeat"
complement(22217..22628)
/note="match: ESTs R50012 R70180"
24456..24530
/note="MIR repeat: matches 41..123 of consensus"
24535..24834
/note="AluX repeat: matches 302..1 of consensus"
24840..25118
/note="L1MC2 repeat: matches 1065..772 of consensus"
25136..25452
/note="AluJb repeat: matches 11..302 of consensus"
25649..25765
/note="MER5A repeat: matches 63..189 of consensus"
26130..26545
/note="MLT1B repeat: matches 1..390 of consensus;
complement(<26546..35701)
/genes="SAPK4"
/note="match: I63675 patent US 5663313"
complement(join(26546..26932,27171..27347,27446..27524,
27782..27861,29276..29347,29456..29570,29695..29742,
30178..30207,30365..30473,33547..33605,34826..34955,
35525..35905))
/genes="SAPK4"
/note="match: cDNAs U93232 U66243 Y10487 AF015256 Y10488
AF004709"
/evidence="not_experimental"
/product="dJ179N16.1 (Stress Activated Protein Kinase 4
(Mitogen Activated Protein Kinase p38delta))"
complement(26546..35905)
/genes="SAPK4"
complement(join(26853..26932,27171..27347,27446..27524,

```

27782. .27861.29276. .29347.29456. .29570.29695. .29742.
30178. .30207.30365. .30473.33547. .33605.34826. .34955.
35525. .35643))
/gene="SAPK4"
/note="match: proteins 014739 015124"
/codon_start=1
/evidence=not_experimental
/product="dj179N16.1 (Stress Activated Protein Kinase 4
(Mitogen Activated Protein Kinase p38delta))"
/protein_id="CAB08438.1"
/db_xref="GI:3036774"
/db_xref="SPTREMBL:O15264"
/translation="MSLRKKGFYKQDVNKTAWELPKTYVSPTHVSGAYGVCASID
KRSGKVAIKLSRPFQSEIFAKRAYRELLLLKHMQHENVIGLLDVTFPASSLRNFYD
FYLVPFWOTDLQKMEFSEEEKIQLYVQMLKGLKYHSAGVVRHLKPGNLAVNE
DCELKDLFGLARHADAEMTGYVTVRWYRAPEVILSWMHYNQTVDINSVGICMAEMLT

Query Match 13.9% Score 19; DB 10; Length 172048;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGATG 19
|||||
DB 41760 CCACCGCACCTGGCTGATG 41742

RESULT 18
AC007158
LOCUS AC007158 204027 bp DNA HTG 26-OCT-1999
DEFINITION Homo sapiens clone hRPK.90_A_1, *** SEQUENCING IN PROGRESS ***, 2
ORDERED pieces.
ACCESSION AC007158
VERSION AC007158.5 GI:5757580
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 204027)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens, clone hRPK.90_A_1
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 204027)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,

Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,

Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heath,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,

Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,

Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,

Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,

Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 22, 1999 this sequence version replaced gi:4726136.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and

* the accession number will be preserved
* 1 160375; contig of 160375 bp in length
* gap of unknown length
* 160376 204027; contig of 43652 bp in length.

FEATURES
source
1..204027
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.90_A_1"
/clone_lib="rPCI-11 Human Male BAC"

BASE COUNT 63998 a 40107 c 39047 g 60875 t
ORIGIN

Query Match 13.9% Score 19; DB 42; Length 204027;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGATG 19
|||||
DB 111376 CCACCGCACCTGGCTGATG 111394

RESULT 19
HUMVIPPHM1

LOCUS HUMVIPPHM1 270 bp DNA PRI 14-JAN-1995
DEFINITION Human vasoactive intestinal peptide and histidine-methionine amide
peptide hormone (VIP/PHM-27) gene, exon 1.

ACCESSION M14618
VERSION M14618.1 GI:340266
KEYWORDS histidine-methionine amide; hormone; peptide hormone; vasoactive
intestinal peptide.

SEGMENT 1 of 6
SOURCE Human DNA (libraries of Lawn and EMBL), clones lambda-VIP-D and
lambda-VIP-4.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 270)
AUTHORS Linder,S., Barkhem,T., Norberg,A., Persson,H., Schalling,M.,
Hokfelt,T. and Magnusson,G.

TITLE Structure and expression of the gene encoding the vasoactive
intestinal peptide precursor

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (2), 605-609 (1987)
MEDLINE 87092456
COMMENT Draft entry and clean copy of sequence [1] kindly provided by
S.Linder, 02-MAR-1987.

FEATURES
source
1..270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6q24-q27"

prim_transcript 100. >270
/gene="vip"
/note="VIP/PHM27 mRNA; G00-120-490"

intron 264. >270
/gene="VIP"
/note="VIP/PHM27 mRNA intron A"

BASE COUNT 76 a 67 c 69 g 58 t
ORIGIN Unreported.

Query Match 13.1% Score 18; DB 9; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TGACTTCTTTCAGAGGAC 47
|||||

DB 22 TGACTTCTTTCAGAGGAC 39
|||||

RESULT 20
G24450

LOCUS G24450 474 bp DNA STS 31-MAY-1996
DEFINITION human STS WI-13939, sequence tagged site.
ACCESSION G24450
VERSION G24450.1 GI:1344776
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS
JOURNAL Unpublished (1995)
COMMENT Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: CTCCTCTTTCTATTACCCAC
Primer B: GTGAGAGAGTCTTCCGAGG
STS size: 150
PCR Profile:
 Presoak:
 Denaturation: 56 degrees C
 Annealing: 56 degrees C
 Polymerization: 35
 PCR Cycles: 35
 Thermal Cycler:
Protocol:
 Template: 10 ng
 Primers: each 5 pM
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Derived from dbEST (genbank accession T17065).
FEATURES
 source
 1..474
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="441.4 CR from top of Chr17 linkage group"
 1..150
 primer_bind
 1..20
 primer_bind complement(131..150)
BASE COUNT 127 a 117 c 101 g 123 t 6 others
ORIGIN
Query Match 13.1%; Score 18; DB 13; Length 474;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 52 AAAGACACTAAGTGAAT 69
|||||
Db 258 AAAGACACTAAGTGAAT 275

Query Match 13.1%; Score 18; DB 11; Length 5190;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 GCTGATGCTTTCTATCT 30
|||||

LOCUS HSU66615 5190 bp mRNA PRI 18-SEP-1996
DEFINITION Human SWI/SNF complex 155 KDa subunit (BAF155) mRNA, complete cds.

U66615
U66615.1 GI:1549238
human.
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5190)
AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
TITLE Diversity and specialization of mammalian SWI/SNF complexes
JOURNAL Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE 96397413
REFERENCE 2 (bases 1 to 5190)
AUTHORS Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kalpana, G. V., Goff, S. P., Yaniv, M., Workman, J. L. and Crabtree, G. R.
TITLE Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
JOURNAL EMBO J. 15 (1996) In press
REFERENCE 3 (bases 1 to 5190)
AUTHORS Wang, W., Xue, Y., Zhou, S. and Crabtree, G. R.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
FEATURES
 source
 1..5190
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="Jurkat T cells"
 /genes="BAF155"
 /55..3369
 /55..3369
 /genes="BAF155"
 /notes="a core subunit presents in all human SWI/SNF complexes purified so far; similar to subunit BAF170 and yeast Swi3 protein; contains a region similar to DNA binding domain of myb and a predicted leucine zipper; the C-terminus of the protein is highly proline-rich and somewhat glutamine-rich"
 /codon_start=1
 /product="SWI/SNF complex 155 KDa subunit"
 /protein_id="AAC50893.1"
 /db_xref="GI:1549239"
 /translation="MAAAGGGGGTAVGATGFGDSAAAGLAVYRRKDGGPATKFEW
 SPETVSQDSVRVWLGKHYKVHADAPNTKTLAGLVOLLOQFQEDAFGKHVTNPATF
 KLPAKCFMDFKAGALCHILGAAYKYKNEGWRFFDLONPSRMDNRVFMFNTKTLV
 ONNCLTEPNLYLPIDILKLANKLDILKHHOGFTDEKSKASHIYPYSSODDEFW
 LRPMRKEKQVLVHWGFPDPSYDTWHSNDVDAEIEDPPIPEKVPKWHVHWILDTDF
 NEWNEEDYEVDENRRPVPFRISTKNEEPVRSERRRORRKSANARKRHSPSPPPP
 TPTESRKKSGKGAASLYKKRSQKEDEEDLTDMEDPTVPVTEEVYLPKNVNLK
 KDSENTPVKGTVDLDEQDEETVTAGKDEDEDPKQDSRSVDLGEDNVTQTHII
 IPSYASWFYNCIHTERRALPEFFGNKSKTPEIYLAYRNMIDSYRLNPOEYLT
 TACRRNLTGDCVAMRVHAGGEOMGLVNYQVDPESRPMAGPPTPHFNVLADTPLAC
 ASDLRSQVPAACOMLNFPEKKEKQVDFLQNFGLRTDIYSKTKLAKSKASAGRWTE
 QETLLLEALMDNDNNKRVSEHVSRQTQECILHFLRPDIEDPIEDNSDASLGLPLA
 OPVPSQSGNPVMTVAFLASVVDPRVSAAKAALEEFSSRVEEVPVLELVEAHVKV
 QEAARAGKVDPTYGLSESSCIAGTGPDEPEKLEGAEEKEADPDQCGQEKAKNVN
 ETDEGKAQDGENEKSEKQSEVSDTKSEKETEENKELSTCKRESQTKGKKV
 EHEISEGNVATAAALASATKAKHLAAVEERKIKSLVALLVETOMKLEIKLRHFE
 GLETIMDRKEALEQOQOOLLTERONFHEOLKYAELRARQOQEQOHOHQOQAOH
 SGFGPLAIGAGHGMHPOOPFPLMHQMPHPPOGOIPGSGMSPGHPGHPG
 RMPTVAANTHPSGSGTTPPGMPMPGNILGPRVPLTAPNGMTPPPPPQPPPPPAD
 GVPPPPAGPPASAAP"
BASE COUNT 1461 a 1168 c 1297 g 1264 t
ORIGIN

Query Match 13.1%; Score 18; DB 11; Length 5190;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 GCTGATGCTTTCTATCT 30
|||||

Db 1013 GCTGATGCTTTTCTATCT 996

RESULT 22
AC008821/c
LOCUS AC008821 14837 bp DNA HTG 03-AUG-1999
DEFINITION Homo sapiens chromosome 5 clone CITB-HL_2129G21, *** SEQUENCING IN
PROGRESS ***, 18 unordered pieces.
ACCESSION AC008821
VERSION AC008821.1 GI:5686250
KEYWORDS HTG; HIGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 14837)
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14837)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 649: contig of 649 bp in length
* gap of unknown length
* 650 1303: contig of 654 bp in length
* gap of unknown length
* 1304 1962: contig of 659 bp in length
* gap of unknown length
* 1963 2611: contig of 649 bp in length
* gap of unknown length
* 2612 2733: contig of 122 bp in length
* gap of unknown length
* 2734 3378: contig of 645 bp in length
* gap of unknown length
* 3379 4036: contig of 658 bp in length
* gap of unknown length
* 4037 4670: contig of 634 bp in length
* gap of unknown length
* 4671 5597: contig of 927 bp in length
* gap of unknown length
* 5598 6302: contig of 705 bp in length
* gap of unknown length
* 6303 7324: contig of 1022 bp in length
* gap of unknown length
* 7325 7966: contig of 642 bp in length
* gap of unknown length
* 7967 8984: contig of 1018 bp in length
* gap of unknown length
* 8985 9882: contig of 898 bp in length
* gap of unknown length
* 9883 10526: contig of 644 bp in length
* gap of unknown length
* 10527 11913: contig of 1387 bp in length
* gap of unknown length
* 11914 13303: contig of 1390 bp in length
* gap of unknown length
* 13304 14837: contig of 1534 bp in length.
Location/Qualifiers
1..14837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"

BASE COUNT 4433 a 3077 c 3067 g 4254 t 6 others
ORIGIN

Query Match 13.1%; Score 18; DB 41; Length 14837;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACGGCACCCTGGCTGAT 18
|||||
Db 12740 CCACGGCACCCTGGCTGAT 12723

RESULT 23
AC002120/c
LOCUS AC002120 23536 bp DNA PRI 31-JUL-1998
DEFINITION Homo sapiens chromosome 9q34, clone 182E8, complete sequence.
ACCESSION AC002120
VERSION AC002120.1 GI:3366591
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 23536)
TITLE Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 23536)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
Gage,D., Geraghty,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McMurtrei,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J., and
Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
TITLE Submitted (14-MAY-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
REFERENCE 3 (bases 1 to 23536)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatlin,C.,
Butwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nachman,A., Nafz,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye.W.J., Zhao,J. and
Zody,M.
Direct Submission
TITLE Submitted (31-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
COMMENT On Jul 31, 1998 this sequence version replaced gi:3184311.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only 23.5 kilobases of this clone are being submitted. The
remainder overlaps either accession number AC000393 (WICGR project
L174) or accession number AC000394 (WICGR project L177).

Basepairs from 15764 to 23536 are part of a region with a tendency
to delete. These base pairs were determined entirely from direct
sequencing of the clone and PCR products.
Location/Qualifiers

FEATURES

source	1. .23536	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="182E8"	
	/clone_lib="Dr. David Kwiatkowski"	
	/map="9c34"	
	/chromosome="9"	
repeat_region	complement(215. .321)	
repeat_region	/rpt_family="MIR"	
repeat_region	1680. .1774	
repeat_region	/rpt_family="L2"	
repeat_region	1892. .2023	
repeat_region	/rpt_family="AluJo"	
repeat_region	2024. .2341	
repeat_region	/rpt_family="AluSg"	
repeat_region	2414. .2448	
repeat_region	/rpt_family="GAAA)n"	
repeat_region	2754. .2888	
repeat_region	/rpt_family="AluSg/x"	
repeat_region	2890. .3185	
repeat_region	/rpt_family="AluY"	
repeat_region	3188. .3325	
repeat_region	/rpt_family="AluSg/x"	
repeat_region	3493. .3520	
repeat_region	/rpt_family="CA)n"	
repeat_region	4310. .4406	
repeat_region	/rpt_family="(GA)n"	
repeat_region	complement(5200. .5496)	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(5828. .5928)	
repeat_region	/rpt_family="L2"	
repeat_region	5978. .6111	
repeat_region	/rpt_family="AluSg/x"	
repeat_region	6112. .6413	
repeat_region	/rpt_family="AluSp"	
repeat_region	complement(6782. .6812)	
repeat_region	/rpt_family="(GAAA)n"	
repeat_region	complement(6824. .7114)	
repeat_region	/rpt_family="AluJb"	
repeat_region	complement(7117. .7337)	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(8457. .8514)	
repeat_region	/rpt_family="L2"	
repeat_region	10625. .11083	
repeat_region	/rpt_family="L1MB3"	
repeat_region	11366. .11503	
repeat_region	/rpt_family="L2"	
repeat_region	11681. .11862	
repeat_region	/rpt_family="(TGGA)n"	
repeat_region	11878. .11996	
repeat_region	/rpt_family="(TGGA)n"	
repeat_region	11998. .12100	
repeat_region	/rpt_family="(TGGA)n"	
repeat_region	complement(12478. .12512)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(12789. .12913)	
repeat_region	/rpt_family="(GA)n"	
repeat_region	13842. .14006	
repeat_region	/rpt_family="L1MC2"	
repeat_region	complement(14203. .14225)	
repeat_region	/rpt_family="(TA)n"	
repeat_region	14992. .15160	
repeat_region	/rpt_family="L1ME2"	
repeat_region	15187. .15494	
repeat_region	/rpt_family="AluJb"	
repeat_region	15509. .15936	
repeat_region	/rpt_family="L1M4"	
repeat_region	15939. .16234	
repeat_region	/rpt_family="AluSg"	
repeat_region	16236. .16795	
repeat_region	/rpt_family="L1M4"	
repeat_region	16796. .17086	
repeat_region	/rpt_family="AluSg"	
repeat_region	17088. .17250	
repeat_region	/rpt_family="AluJb"	
repeat_region	17277. .17581	
repeat_region	/rpt_family="AluSg"	
repeat_region	17723. .18068	
repeat_region	/rpt_family="L1ME3A"	
repeat_region	18073. .18370	
repeat_region	/rpt_family="AluSg"	
unsure	complement(18154. .18174)	
repeat_region	/note="Single-stranded coverage."	
repeat_region	18372. .18638	
repeat_region	/rpt_family="AluJb"	
repeat_region	18643. .19029	
repeat_region	/rpt_family="L1ME3A"	
repeat_region	complement(19050. .19522)	
unsure	/rpt_family="MLTIG"	
unsure	19077. .19167	
unsure	/note="Single-stranded coverage."	
unsure	complement(19559. .19580)	
unsure	/note="Single-stranded coverage."	
repeat_region	19632. .19779	
repeat_region	/rpt_family="MLTIA2"	
repeat_region	19781. .19878	
repeat_region	/rpt_family="AluSg/x"	
repeat_region	19888. .20108	
repeat_region	/rpt_family="MLTIA2"	
repeat_region	20109. .20435	
repeat_region	/rpt_family="L1ME3A"	
repeat_region	20436. .20736	
repeat_region	/rpt_family="AluSg"	
repeat_region	20738. .20795	
repeat_region	/rpt_family="L1ME3"	
repeat_region	20805. .21205	
repeat_region	/rpt_family="L1M4"	
unsure	complement(21059. .21207)	
unsure	/note="Single-stranded terminator coverage."	
repeat_region	complement(21284. .21555)	
repeat_region	/rpt_family="AluSg"	
repeat_region	21579. .22388	
repeat_region	/rpt_family="TIGGER1"	
unsure	21582	
unsure	/note="Uncertain number of A's; either 24, 25, or 26 A's."	
unsure	complement(22159. .22762)	
unsure	/note="Single-stranded terminator coverage."	
repeat_region	22389. .22685	
repeat_region	/rpt_family="AluSp"	
repeat_region	22686. .22994	
repeat_region	/rpt_family="TIGGER1"	
repeat_region	22996. .23288	
repeat_region	/rpt_family="AluJb"	
repeat_region	23291. .23458	
repeat_region	/rpt_family="TIGGER1"	
repeat_region	complement(23460. .23493)	
repeat_region	/rpt_family="L1MB3"	
repeat_region	complement(23500. .23535)	
repeat_region	/rpt_family="Alu"	
BASE COUNT	6457 a 5475 c 6200 g 5404 t	
ORIGIN		
Query Match	13.1%; Score 18; DB 11; Length 23536;	
Best Local Similarity	100.0%; Pred. No. 6;	
Matches	18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 CCACCGACCTGGCTGAT 18	
Db	23009 CCACCGACCTGGCTGAT 22992	
RESULT	24	
U73643/c	33602 bp DNA	
LOCUS	U73643	
DEFINITION	Human Chromosome 11 Cosmid cSRL34e5, complete sequence.	
	PRI	24-JUL-1997

```

ACCESSION      U73643
VERSION        U73643.1  GI:2276400
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
REFERENCE
AUTHORS      Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
              Davies,J., Davies,C.J., Davis,C., English,C., Fondon,T.,
              Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
              Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U.,
              Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
              Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
              Ward,T. and Wilson,R.
TITLE        Template
JOURNAL
JOURNAL      Unpublished
REFERENCE
AUTHORS      Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davies,J.,
              Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
              Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
              Hinson,S., McFarland,J., Megarity,C., Narayanaswamy,U., Newton,J.,
              O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
              Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
              Ward,T. and Wilson,R.
TITLE        Direct Submission
JOURNAL
JOURNAL      Submitted (07-OCT-1996) McDermott Center for Human Growth and
              Development, University of Texas Southwestern Medical Center at
              Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE
AUTHORS      Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
              Davies,J., Davies,C.J., Davis,C., English,C., Fondon,T.,
              Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
              Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U.,
              Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
              Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
              Ward,T. and Wilson,R.
TITLE        Direct Submission
JOURNAL
JOURNAL      Submitted (24-JUL-1997) Genome Science and Technology Center,
              University of Texas Southwestern Medical Center at Dallas, 5323
              Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT      On Jul 24, 1997 this sequence version replaced gi:1737198.
FEATURES
SOURCE      1..33602
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="CSRL34e5"
              /chromosome="11"
              complement(448..727)
              /rpt_family="Alu"
              1040..1365
              /rpt_family="Alu"
              complement(2027..2078)
              /rpt_family="Alu"
              complement(2289..2841)
              /rpt_family="Alu"
              complement(2890..3084)
              /rpt_family="Alu"
              complement(3654..3938)
              /rpt_family="Alu"
              complement(4104..4350)
              /rpt_family="Alu"
              5820..6110
              /rpt_family="Alu"
              6336..6591
              /rpt_family="Alu"
              6852..7070
              /rpt_family="MIR"
              complement(7083..7358)
              /rpt_family="Alu"
              complement(7621..7793)
              /rpt_family="MIR"
              8519..8779

repeat_region      /rpt_family="Alu"
                    9325..9613
repeat_region      /rpt_family="Alu"
                    10632..10680
repeat_region      /rpt_family="MIR"
                    complement(10811..11084)
repeat_region      /rpt_family="Alu"
                    11827..12092
repeat_region      /rpt_family="Alu"
                    complement(12178..12249)
repeat_region      /rpt_family="MIR"
                    complement(13207..13495)
repeat_region      /rpt_family="Alu"
                    complement(13508..14101)
repeat_region      /rpt_family="Alu"
                    14590..14772
repeat_region      /rpt_family="MIR"
                    16109..16401
repeat_region      /rpt_family="Alu"
                    complement(18386..18965)
repeat_region      /rpt_family="Alu"
                    complement(19286..19873)
repeat_region      /rpt_family="LTR9"
                    19966..20261
repeat_region      /rpt_family="Alu"
                    complement(21464..21765)
repeat_region      /rpt_family="Alu"
                    complement(22974..23553)
repeat_region      /rpt_family="Alu"
                    complement(24905..25021)
repeat_region      /rpt_family="MIR"
                    complement(25246..25504)
repeat_region      /rpt_family="Alu"
                    28658..28950
repeat_region      /rpt_family="Alu"
                    complement(29341..29831)
repeat_region      /rpt_family="L1"
                    30332..30627
repeat_region      /rpt_family="Alu"
                    complement(31736..32027)
repeat_region      /rpt_family="Alu"
                    complement(32818..32919)
repeat_region      /rpt_family="Alu"

BASE COUNT      8284 a 8121 c 8156 g 9041 t
ORIGIN
Query Match      13.1%; Score 18; DB 10; Length 33602;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CCACCGCACCTGGCTGAT 18
          |||||
Db      28664 CCACCGCACCTGGCTGAT 28647

RESULT      25
AC014889/c
LOCUS      AC014889
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
              pieces.
ACCESSION  AC014889
VERSION    AC014889.1  GI:6436446
KEYWORDS   HTG: HTGS_PHASE2.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 35967)
AUTHORS   Adams,M. and Venter,J.C.
TITLE     Direct Submission
JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

```



```

Rockville, MD, USA
This sequence was identified as CDM:10212742 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source
        Location/Qualifiers
            1..35967
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
BASE COUNT  10857 a  7230 c  7235 g  10645 t
ORIGIN
Query Match      13.1%; Score 18; DB 43; Length 35967;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  82  TCTTCCAAGCTAAACAA 99
      |||||
Db  24650 TCTTCCAAGCTAAACAA 24633

RESULT 26
HS06G2B/c
LOCUS      HS06G2B      37027 bp      DNA      PRI      23-NOV-1999
DEFINITION Human DNA sequence from PAC 506G2 contains STSS and a CpG island.
ACCESSION 282976
VERSION   282976.1  GI:1673498
KEYWORDS  CpG island; X.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 37027)
           Deadman.R.
           Direct Submission
           Submitted (13-NOV-1996) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1RO, UK. E-mail enquiries: humquery@sanger.ac.uk
           requests: clonerequest@sanger.ac.uk
           IMPORTANT: This sequence is not the entire insert of clone 506G2.
           It may be shorter because we only sequence overlapping sections
           once, or longer because we arrange for a small overlap between
           neighbouring submissions.
           This sequence has been finished according to sequence map criteria
           as follows. An attempt is made to resolve all sequencing problems,
           such as compressions and repeats, but not necessarily within known
           annotated human repeat sequence elements (e.g. Alu). Where the
           sequence is ambiguous, there is an annotation using the 'unsure'
           feature key.
           The sequence from clone 506G2 has been finished in more than one
           contig. This sequence (506G2B) is separated from the preceding one
           (506G2A) by a gap of 200bp sized by PCR and restriction analysis.
           The true right end of clone 506G2 is at 37027.
           506G2 is from a whole genomic PAC library.
           Location/Qualifiers
               1..37027
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="X"
                   /map="X"
                   /clone="RP3-506G2"
                   /clone_lib="RPCI-3"
                   1..125
                   /partial
                   /notes="AluSg repeat: matches 125. .1 of consensus"
                   163..281
                   /partial
                   /note="AluJo repeat: matches 11. .127 of consensus"
                   296..596
                   /note="AluSp repeat: matches 1. .302 of consensus"
                   604..772
                   /note="AluSg repeat: matches 1. .93 of consensus"

```

```
repeat_region 13651..13919
/partial
/note="AluJb repeat: matches 60. .301 of consensus"
repeat_region 14110..14432
/note="AluJb repeat: matches 13. .302 of consensus"
repeat_region 14722..15024
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 15222..15326
/partial
/note="AluJb repeat: matches 196. .300 of consensus"
repeat_region 15714..16016
/note="AluJb repeat: matches 1. .290 of consensus"
repeat_region 16424..16531
/note="FLAN_A repeat: matches 117. .6 of consensus"
repeat_region 18974..19272
/note="AluSp repeat: matches 303. .2 of consensus"
repeat_region 19497..19756
/partial
/note="AluY repeat: matches 39. .298 of consensus"
repeat_region 20312..20478
/note="MIR repeat: matches 260. .69 of consensus"
misc_feature 21146..23225
/note="Putative CpG island"
repeat_region 22691..23782
/note="SVA repeat: matches 1101. .1 of consensus"
repeat_region 23783..23828
/note="23 copies of 2 mer 80 % conserved"
repeat_region 24117..24481
/note="L1PA5 repeat: matches 517. .890 of consensus"
repeat_region 25101..25284
/note="MER20 repeat: matches 155. .1 of consensus"
repeat_region 26018..26081
/note="32 copies of 2 mer 83 % conserved"
repeat_region 26083..26258
/partial
/note="AluJb repeat: matches 195. .5 of consensus"
repeat_region 26463..26760
/note="AluY repeat: matches 3. .300 of consensus"
repeat_region 26761..27058
/note="AluSp repeat: matches 2. .302 of consensus"
repeat_region 27676..27988
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region 28413..28685
/note="AluX repeat: matches 298. .2 of consensus"
repeat_region 28718..28848
/note="L1 repeat: matches 3151. .3280 of consensus"
repeat_region 29553..29796
/note="MLT2_internal repeat: matches 5375. .5124 of consensus"
repeat_region 30109..30302
/note="MLT2_internal repeat: matches 4754. .4572 of consensus"
repeat_region 30324..30621
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 30655..30952
/note="AluSg repeat: matches 1. .296 of consensus"
repeat_region 30955..31101
/note="MER4B repeat: matches 11. .148 of consensus"
repeat_region 31637..31915
/note="AluX repeat: matches 1. .287 of consensus"
repeat_region 31977..33195
/note="MLT2_internal repeat: matches 4572. .3316 of consensus"
repeat_region 33199..33502
/note="AluJo repeat: matches 1. .302 of consensus"
repeat_region 33616..33917
/note="AluJb repeat: matches 1. .301 of consensus"
repeat_region 33984..34529
/note="MLT2_internal repeat: matches 3168. .2597 of consensus"
repeat_region 34644..34800
/note="MLT2_internal repeat: matches 2419. .2264 of consensus"
```

```
repeat_region 35043..35343
/note="AluY repeat: matches 1. .301 of consensus"
misc_feature 35348..35643
/note="match:STS L42688"
repeat_region 35350..35569
/note="MLT2_internal repeat: matches 1972. .1754 of consensus"
Query Match 13.1%; Score 18; DB 11; Length 37027;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 4309 CCACCGCACCTGGCTGAT 4292
|||||

RESULT 27
AC000394/c
LOCUS AC000394 44294 bp DNA PRZ 29-MAY-1997
DEFINITION Genomic sequence from Human 9q34, complete sequence.
AC000394
AC000394
AC000394.1 GI:2133911
VERSION HTG
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 44294)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H. and Lander,E.S.
TITLE Genomic sequence from Human 9q34
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 44294)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (08-APR-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 44294)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 29, 1997 this sequence version replaced gi:1929461.
The Staden databases, finishing information, and all
chromatographic files used in the assembly of this clone are
available from our anonymous ftp site.
All repeats were identified using RepeatMasker: Smit, A.F.A. 6
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1. 44294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="81F1"
/clone_lib="Dr. David Kwiatkowski"
/chromosome="9q34"
repeat_region 1. .244
```

```
repeat_region /rpt_family="TIGGER1"
repeat_region 246. .538
repeat_region /rpt_family="AluJb"
repeat_region 541. .708
repeat_region /rpt_family="TIGGER1"
repeat_region complement(693. .743)
repeat_region /rpt_family="L1MB3"
repeat_region complement(750. .1045)
repeat_region /rpt_family="AluSx"
repeat_region complement(1054. .1183)
repeat_region /rpt_family="AluJo"
repeat_region 1211. .1502
repeat_region /rpt_family="AluSx"
repeat_region 3258. .3300
repeat_region /rpt_family="(TAA)n"
repeat_region complement(3306. .3361)
repeat_region /rpt_family="L1MB8"
repeat_region 3713. .4016
repeat_region /rpt_family="AluSx"
repeat_region 4017. .4062
repeat_region /rpt_family="POLY-A"
repeat_region complement(4233. .4254)
repeat_region /rpt_family="AT-rich"
repeat_region complement(4832. .5135)
repeat_region /rpt_family="AluSx"
repeat_region 5744. .5865
repeat_region /rpt_family="MIR2"
repeat_region complement(5892. .6211)
repeat_region /rpt_family="MLT1A1"
repeat_region 6460. .6759
repeat_region /rpt_family="AluSg"
repeat_region complement(6826. .6947)
repeat_region /rpt_family="MLT1B"
repeat_region 7299. .8285
repeat_region /rpt_family="LTR13"
repeat_region complement(8500. .8858)
repeat_region /rpt_family="MSTD"
repeat_region complement(8968. .9269)
repeat_region /rpt_family="AluSx"
repeat_region complement(9566. .9705)
repeat_region /rpt_family="MIR2"
repeat_region 10801. .11099
repeat_region /rpt_family="AluSc"
repeat_region 11111. .11319
repeat_region /rpt_family="AluJo"
repeat_region complement(111377. .11634)
repeat_region /rpt_family="AluJo"
repeat_region 11817. .12107
repeat_region /rpt_family="AluSx"
repeat_region 12293. .12334
repeat_region /rpt_family="MIR2"
repeat_region 13306. .13462
repeat_region /rpt_family="LTR10"
repeat_region complement(13467. .13688)
repeat_region /rpt_family="AluSg"
repeat_region 13689. .13988
repeat_region /rpt_family="LTR10"
repeat_region complement(14257. .14347)
repeat_region /rpt_family="AluSg"
repeat_region 14349. .14514
repeat_region /rpt_family="LTR10"
repeat_region 20090. .20214
repeat_region /rpt_family="LTR10"
repeat_region complement(20571. .20875)
repeat_region /rpt_family="AluSx"
repeat_region complement(20980. .21025)
repeat_region /rpt_family="MIR"
repeat_region complement(21084. .21201)
repeat_region /rpt_family="MIR"
repeat_region 21438. .21513
repeat_region /rpt_family="MIR2"
repeat_region complement(21522. .21805)
repeat_region /rpt_family="AluJo"
```

```
repeat_region complement(21912. .21978)
repeat_region /rpt_family="MIR"
repeat_region complement(22248. .22295)
repeat_region /rpt_family="MIR2"
repeat_region 22898. .23262
repeat_region /rpt_family="LTR13"
repeat_region 23261. .23894
repeat_region /rpt_family="LTR13"
repeat_region 24976. .25119
repeat_region /rpt_family="MIR"
repeat_region 25172. .25264
repeat_region /rpt_family="(GAA)n"
repeat_region 25425. .25480
repeat_region /rpt_family="MERSA"
repeat_region complement(25534. .25577)
repeat_region /rpt_family="MADE1"
repeat_region 26120. .26286
repeat_region /rpt_family="AluSg"
repeat_region 27580. .27687
repeat_region /rpt_family="MIR2"
repeat_region complement(27754. .28057)
repeat_region /rpt_family="AluSx"
repeat_region 28165. .28468
repeat_region /rpt_family="AluJo"
repeat_region complement(29642. .29942)
repeat_region /rpt_family="AluSg"
repeat_region 29964. .30260
repeat_region /rpt_family="AluJb"
repeat_region 30337. .30637
repeat_region /rpt_family="AluY"
repeat_region 30638. .30803
repeat_region /rpt_family="AluSp"
repeat_region complement(31900. .32128)
repeat_region /rpt_family="L1MC1"
repeat_region 32269. .32425
repeat_region /rpt_family="MIR"
repeat_region complement(34398. .34533)
repeat_region /rpt_family="(GGAA)n"
repeat_region complement(34952. .35259)
repeat_region /rpt_family="AluSx"
repeat_region complement(35342. .35417)
repeat_region /rpt_family="MIR"
repeat_region 35424. .35718
repeat_region /rpt_family="AluJb"
repeat_region complement(35731. .35865)
repeat_region /rpt_family="MIR"
repeat_region 39994. .40040
repeat_region /rpt_family="(GGA)n"
repeat_region complement(40895. .40738)
repeat_region /rpt_family="(TGAA)n"
repeat_region complement(41319. .41601)
repeat_region /rpt_family="AluJb"
repeat_region complement(41661. .41736)
repeat_region /rpt_family="MER42C"
repeat_region complement(41929. .42224)
repeat_region /rpt_family="MER42C"
repeat_region 42271. .42568
repeat_region /rpt_family="AluJb"
repeat_region 42574. .42745
repeat_region /rpt_family="L1MC1"
repeat_region complement(42768. .42853)
repeat_region /rpt_family="(GGAA)n"

BASE COUNT 10447 a 11319 c 11448 g 11080 t
ORIGIN
```

Query Match 13.1%; Score 18; DB 10; Length 44294;
Best Local Similarity 100.0%; Pred.No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCCGACCTGGCTGAT 18

|||||

DB 259 CCACCCGACCTGGCTGAT 242

RESULT	28
AC012318/c	
LOCUS	AC012318 48925 bp DNA HTG 31-OCT-1999
DEFINITION	Homo sapiens chromosome 19 clone LLNL-F_143B12, *** SEQUENCING IN PROGRESS ***, 23 unordered pieces.
ACCESSION	AC012318
VERSION	AC012318.2 GI:6165034
KEYWORDS	HTG: HTGS_PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 48925)
JOURNAL	DOE Joint Genome Institute.
REFERENCE	Sequencing of Human Chromosome 19
AUTHORS	Unpublished
TITLE	2 (bases 1 to 48925)
JOURNAL	DOE Joint Genome Institute. Direct Submission
COMMENT	Submitted (23-Oct-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Oct 31, 1999 this sequence version replaced gi:6102621. www.jgi.doe.gov . * NOTE: This is a 'working draft' sequence. It currently * consists of 23 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 700: contig of 700 bp in length * gap of unknown length * 701 3126: contig of 2426 bp in length * gap of unknown length * 3127 3577: contig of 451 bp in length * gap of unknown length * 3578 4292: contig of 715 bp in length * gap of unknown length * 4293 5134: contig of 842 bp in length * gap of unknown length * 5135 5766: contig of 632 bp in length * gap of unknown length * 5767 6491: contig of 725 bp in length * gap of unknown length * 6492 7211: contig of 720 bp in length * gap of unknown length * 7212 8169: contig of 958 bp in length * gap of unknown length * 8170 8934: contig of 765 bp in length * gap of unknown length * 8935 9750: contig of 816 bp in length * gap of unknown length * 9751 11887: contig of 2137 bp in length * gap of unknown length * 11888 13847: contig of 1960 bp in length * gap of unknown length * 13848 15721: contig of 1874 bp in length * gap of unknown length * 15722 17899: contig of 2178 bp in length * gap of unknown length * 17900 19846: contig of 1947 bp in length * gap of unknown length * 19847 21938: contig of 2092 bp in length * gap of unknown length * 21939 24668: contig of 2730 bp in length * gap of unknown length * 24669 28196: contig of 3528 bp in length * gap of unknown length * 28197 30957: contig of 2761 bp in length * gap of unknown length *

```

*      *      1785:  contig of 892 bp in length
*      *      gap of unknown length
*      *      2662:  contig of 877 bp in length
*      *      gap of unknown length
*      *      3517:  contig of 855 bp in length
*      *      gap of unknown length
*      *      3518:  contig of 751 bp in length
*      *      gap of unknown length
*      *      4269:  contig of 885 bp in length
*      *      gap of unknown length
*      *      5154:  contig of 904 bp in length
*      *      gap of unknown length
*      *      6058:  contig of 895 bp in length
*      *      gap of unknown length
*      *      6953:  contig of 863 bp in length
*      *      gap of unknown length
*      *      7816:  contig of 880 bp in length
*      *      gap of unknown length
*      *      8696:  contig of 914 bp in length
*      *      gap of unknown length
*      *      9610:  contig of 893 bp in length
*      *      gap of unknown length
*      *      10503: contig of 890 bp in length
*      *      gap of unknown length
*      *      11393: contig of 882 bp in length
*      *      gap of unknown length
*      *      12275: contig of 756 bp in length
*      *      gap of unknown length
*      *      13031: contig of 879 bp in length
*      *      gap of unknown length
*      *      13910: contig of 859 bp in length
*      *      gap of unknown length
*      *      14769: contig of 876 bp in length
*      *      gap of unknown length
*      *      15645: contig of 922 bp in length
*      *      gap of unknown length
*      *      16567: contig of 880 bp in length
*      *      gap of unknown length
*      *      17447: contig of 893 bp in length
*      *      gap of unknown length
*      *      18340: contig of 897 bp in length
*      *      gap of unknown length
*      *      19237: contig of 919 bp in length
*      *      gap of unknown length
*      *      20156: contig of 926 bp in length
*      *      gap of unknown length
*      *      21082: contig of 868 bp in length
*      *      gap of unknown length
*      *      21950: contig of 930 bp in length
*      *      gap of unknown length
*      *      22880: contig of 854 bp in length
*      *      gap of unknown length
*      *      23734: contig of 875 bp in length
*      *      gap of unknown length
*      *      24609: contig of 883 bp in length
*      *      gap of unknown length
*      *      25492: contig of 887 bp in length
*      *      gap of unknown length
*      *      26379: contig of 879 bp in length
*      *      gap of unknown length
*      *      27258: contig of 887 bp in length
*      *      gap of unknown length
*      *      28145: contig of 896 bp in length
*      *      gap of unknown length
*      *      29041: contig of 864 bp in length
*      *      gap of unknown length
*      *      29905: contig of 883 bp in length
*      *      gap of unknown length
*      *      30788: contig of 874 bp in length
*      *      gap of unknown length
*      *      31662: contig of 933 bp in length
*      *      gap of unknown length
*      *      32595: contig of 885 bp in length

```

```

*      *      33480: gap of unknown length
*      *      34403: contig of 924 bp in length
*      *      gap of unknown length
*      *      34404: contig of 868 bp in length
*      *      gap of unknown length
*      *      35272: contig of 890 bp in length
*      *      gap of unknown length
*      *      36162: contig of 872 bp in length
*      *      gap of unknown length
*      *      37034: contig of 892 bp in length
*      *      gap of unknown length
*      *      37926: contig of 899 bp in length
*      *      gap of unknown length
*      *      38825: contig of 864 bp in length
*      *      gap of unknown length
*      *      39689: contig of 887 bp in length
*      *      gap of unknown length
*      *      40576: contig of 891 bp in length
*      *      gap of unknown length
*      *      41467: contig of 925 bp in length
*      *      gap of unknown length
*      *      42392: contig of 875 bp in length
*      *      gap of unknown length
*      *      43267: contig of 865 bp in length
*      *      gap of unknown length
*      *      44132: contig of 875 bp in length
*      *      gap of unknown length
*      *      45007: contig of 914 bp in length
*      *      gap of unknown length
*      *      45921: contig of 881 bp in length
*      *      gap of unknown length
*      *      46802: contig of 881 bp in length
*      *      gap of unknown length
*      *      47683: contig of 891 bp in length
*      *      gap of unknown length
*      *      48574: contig of 893 bp in length
*      *      gap of unknown length
*      *      49467: contig of 911 bp in length
*      *      gap of unknown length
*      *      50378: contig of 873 bp in length
*      *      gap of unknown length
*      *      51251: contig of 887 bp in length
*      *      gap of unknown length
*      *      52138: contig of 938 bp in length
*      *      gap of unknown length
*      *      53076: contig of 856 bp in length
*      *      gap of unknown length
*      *      53932: contig of 891 bp in length
*      *      gap of unknown length
*      *      54823: contig of 885 bp in length
*      *      gap of unknown length
*      *      55707: Location/Qualifiers
*      *      1. .55707
*      *      /organism="Homo sapiens"
*      *      /db_xref="taxon:9606"
*      *      /clone="115_K14"
*      *      /clone_lib="RPC1-11 Human Maie BAC"
*      *      16397 a 10660 c 10769 g 17182 t 699 others
*      *      BASE COUNT
*      *      ORIGIN

```

```

Query Match      13.1%; Score 18; DB 41; Length 55707;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CCACCGACCTGGCTGAT 18
Db      51596 CCACCGACCTGGCTGAT 51579

RESULT 30
LOCUS   AB026647      81542 bp      DNA      PLN      07-MAY-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MJL12,

```

ACCESSION	AB026647	complete sequence.	
VERSION	AB026647.1	GI:4757403	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:WJL12		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (sites)		
AUTHORS	Nakamura,K.		
TITLE	Structural Analysis of Arabidopsis thaliana Chromosome 3. II		
JOURNAL	Unpublished (1999)		
REFERENCE	2 (bases 1 to 81542)		
AUTHORS	Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935(ex.7443), Fax:81-438-52-3934)		
FEATURES	Location/Qualifiers		
source	1. .81542		
	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="3"		
	/clone="WJL12"		
BASE COUNT	25859 a 14006 c 14834 g 26843 t		
ORIGIN			
	13.1%; Score 18; DB 7; Length 81542;		
	Best Local Similarity 100.0%; Pred. No. 5.8;		
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	74 CCTTGAAGTCTTCCAGC 91		
Db	19284 CCTTGAAGTCTTCCAGC 19301		
RESULT 31			
HS931E15/c			
LOCUS	HS931E15 81874 bp DNA PRI 23-NOV-1999		
DEFINITION	Human DNA sequence from clone 931E15 on Chromosome Xq25. Contains STSs, GSSs and genomic marker DXS8098, complete sequence.		
ACCESSION	AL023575		
VERSION	AL023575.1		
KEYWORDS	HTG; DXS8098.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 81874)		
AUTHORS	Pavitt,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	On Sep 18, 1998 this sequence version replaced gi:3550203. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the		


```

repeat_region 26514..26804
/rpt_family="AluJb"
complement(26831..27035)
/rpt_family="MLT1D"
complement(27042..27333)
/rpt_family="AluY"
complement(27334..27637)
/rpt_family="AluSp"
complement(27644..27716)
/rpt_family="MLT1D"
27727..28020
/rpt_family="AluX"
28881..28972
/rpt_family="AluX"
29177..29471
/rpt_family="7SLRNA"
complement(29214..29339)
/notes="predicted exon, program:
frame: 0, quality: good, score: 50.000"
complement(29628..29915)

```

```

Query Match 13.18; Score 18; DB 11; Length 86765;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 19053 CCACCGCACCTGGCTGAT 19036

```

RESULT 33

```

LOCUS HS398C22 89328 bp DNA PRI 23-NOV-1999
DEFINITION Homo sapiens DNA sequence from PAC 398C22 on chromosome
22q11.2-qter. Contains Brain Protein E46 like sequences, ESTs.
ACCESSION 293784
VERSION 293784.1 GI:2315175
KEYWORDS 22q11.2-qter; Brain Protein E46-like.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 89328)
Burgess,J.

```

```

Direct Submission
Submitted (31-JUL-1997) Chromosome 22 Project Group
(http://www.sanger.ac.uk/HGP/Chr22/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 7, 1997 this sequence version replaced gi:1929050.
IMPORTANT: This sequence is not the entire insert of clone 398C22.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre chromosome 22
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true right end of clone 941F9 is at 104.
The true right end of clone 398C22 is at 89328.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
398C22 is from the library RCI3 constructed at the Roswell Park

```

Cancer Institute by the group of Pieter de Jong.
For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES

source

```

1..89328
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2-qter"
/clone="RP3-398C22"
/clone_lib="RPCI-3"

```

```

511..575
/notes="AluJb repeat: matches 296..232 of consensus;
incomplete repeat"
577..735
/notes="AluJb repeat: matches 153..3 of consensus;
incomplete repeat"

```

```

repeat_region 2291..2582
/notes="AluJb repeat: matches 300..2 of consensus"
repeat_region 3011..3311
/notes="AluSx repeat: matches 1..301 of consensus"
repeat_region 3624..3763
/notes="L1 repeat: matches 5042..4898 of consensus"
repeat_region 3924..4226
/notes="AluJb repeat: matches 302..1 of consensus"
repeat_region 6512..6809
/notes="AluSg repeat: matches 1..297 of consensus"
repeat_region 7753..8041
/notes="AluSg repeat: matches 1..290 of consensus"
repeat_region 9331..9533
/notes="MIR repeat: matches 38..251 of consensus"
9587..60512
/genes="DJ398C22.1"

```

```

join(<9687..9877,12970..13052,20256..20352,22663..22821,
39387..38467,49399..49564,58705..58813,60343..>60512)
/genes="DJ398C22.1"
/notes="E46-like contains exons 2-9 continues in 284478"
/codon_start=1
/product="DJ398C22.1 (novel protein, ortholog of mouse
brain protein E46)"
/protein_id="CAB07856.1"
/db_xref="GI:2956661"

```

```

/db_xref="SPTREMBL:O14998"
/translation="ETAPRTIFQVRVLDLKKSHAVELACRDPQSQVENLASSLQLTIE
CFCLRNACIECVNONSIRNLDITIGVADLILLFRELVEQSLTAFRCGLQFLGN
IASRDNDSQSIYVWHPFELFSLCNHPDKKIYVSSMILFTSLNHERKELRENLI
AIDVIDAYQKHPESEWPFIIITDLFLKSPELVQAMPKLNQERVTLIDLMIAKITSD
EPLTKDDIIFVFLRHAEILASTFVDQCKTVLKLASEPDPDEALATIRLDVLCEKTV
NTELLGYLOVFELCLLRLVIRVHACKETNIFSNCGCVRAEGDISNVANGFKSH
LRLIGNLCYKKNQDNQK"
9965..10269
/notes="AluJb repeat: matches 1..302 of consensus"
11343..11529
/notes="MIR repeat: matches 223..30 of consensus"
11588..11877
/notes="AluSg repeat: matches 1..291 of consensus"
11942..12099
/notes="MIR repeat: matches 86..259 of consensus"
13267..133632
/notes="MT1A1 repeat: matches 358..1 of consensus"
13635..13927
/notes="MT1B repeat: matches 44..358 of consensus"
13947..14046
/notes="MIR2 repeat: matches 109..15 of consensus"
14344..14473
/notes="AluJb repeat: matches 1..131 of consensus;
incomplete repeat"
14474..14770
/notes="AluSp repeat: matches 1..297 of consensus"
14771..14934
/notes="AluSx repeat: matches 138..302 of consensus;
incomplete repeat"
15023..15315
/notes="AluY repeat: matches 1..298 of consensus"

```

```

repeat_region 9965..10269
/notes="AluJb repeat: matches 1..302 of consensus"
repeat_region 11343..11529
/notes="MIR repeat: matches 223..30 of consensus"
repeat_region 11588..11877
/notes="AluSg repeat: matches 1..291 of consensus"
repeat_region 11942..12099
/notes="MIR repeat: matches 86..259 of consensus"
repeat_region 13267..133632
/notes="MT1A1 repeat: matches 358..1 of consensus"
repeat_region 13635..13927
/notes="MT1B repeat: matches 44..358 of consensus"
repeat_region 13947..14046
/notes="MIR2 repeat: matches 109..15 of consensus"
repeat_region 14344..14473
/notes="AluJb repeat: matches 1..131 of consensus;
incomplete repeat"
repeat_region 14474..14770
/notes="AluSp repeat: matches 1..297 of consensus"
repeat_region 14771..14934
/notes="AluSx repeat: matches 138..302 of consensus;
incomplete repeat"
repeat_region 15023..15315
/notes="AluY repeat: matches 1..298 of consensus"

```

```
repeat_region 15465..15765
/note="Alusg repeat: matches 297. .1 of consensus"
prim_transcript 16256..17096
/gene="dJ398C22.1"
/note="match: multiple ESTs; match: R10652 R99219 R99220
R10653"
repeat_region 17008..17105
/note="L1MB4 repeat: matches 904. .809 of consensus"
repeat_region 17116..17408
/note="Alusx repeat: matches 294. .1 of consensus"
repeat_region 17414..17502
/note="L1MB5 repeat: matches 795. .608 of consensus"
repeat_region 17605..17903
/note="Aluy repeat: matches 300. .1 of consensus"
repeat_region 17905..18486
/note="L1MB8 repeat: matches 622. .4 of consensus"
repeat_region 18489..18603
/note="L1 repeat: matches 4584..4471 of consensus"
repeat_region 19465..19765
/note="Alusx repeat: matches 301. .3 of consensus"
repeat_region 19822..20085
/note="Alusx repeat: matches 299. .1 of consensus"
repeat_region 20399..20661
/note="L1MB2 repeat: matches 654. .911 of consensus"
repeat_region 20818..20913
/note="3 copies of 32 mer 85 & conserved"
repeat_region 24126..25411
/note="L1 repeat: matches 4782. .3450 of consensus"
repeat_region 25415..25707
/note="Alusg repeat: matches 291. .1 of consensus"
repeat_region 25712..25797
/note="L1 repeat: matches 3463. .3375 of consensus"
repeat_region 25800..26146
/note="MER1B repeat: matches 1. .337 of consensus"
repeat_region 26148..27122
/note="L1 repeat: matches 3379. .2343 of consensus"
repeat_region 28621..28908
/note="Alusx repeat: matches 293. .5 of consensus"
repeat_region 30498..30588
/note="MIR2 repeat: matches 1. .96 of consensus"
repeat_region 31144..31683
/note="L1MB2 repeat: matches 545. .1088 of consensus"
repeat_region 33748..34075
/note="L1MB1 repeat: matches 633. .290 of consensus"
repeat_region 34358..34655
/note="AluJo repeat: matches 301. .1 of consensus"
repeat_region 35400..35616
/note="MER30 repeat: matches 230. .1 of consensus"
repeat_region 35953..36259
/note="AluJo repeat: matches 2. .302 of consensus"
repeat_region 36300..36610
/note="Alusx repeat: matches 301. .9 of consensus"
repeat_region 39850..39964
/note="MIR2 repeat: matches 5. .125 of consensus"
repeat_region 40247..40290
/note="L1MB3 repeat: matches 923. .880 of consensus"
repeat_region 40291..40470
/note="FAM repeat: matches 174. .5 of consensus"
repeat_region 40475..40598
/note="L1MB2 repeat: matches 880. .757 of consensus"
repeat_region 40597..40744
/note="L1MB8 repeat: matches 176. .314 of consensus"
repeat_region 40744..40918
/note="MSTD repeat: matches 394. .217 of consensus"
repeat_region 41353..41434
/note="MIR repeat: matches 145. .234 of consensus"
repeat_region 41694..41761
/note="MIR repeat: matches 191. .119 of consensus"
repeat_region 41762..41907
/note="Alusg repeat: matches 154. .299 of consensus;
incomplete repeat"
repeat_region 43244..43313
/note="2 copies of 35 mer 99 & conserved"
```

```
repeat_region 43315..43384
/note="2 copies of 35 mer 91 & conserved"
repeat_region 43385..43649
/note="L1PB2 repeat: matches 874. .604 of consensus"
repeat_region 43650..43944
/note="Alusx repeat: matches 3. .301 of consensus"
repeat_region 43949..44542
/note="L1PB1 repeat: matches 608. .1 of consensus"
repeat_region 44395..46659
/note="L1 repeat: matches 5390. .3097 of consensus"
repeat_region 46669..46954
/note="AluJb repeat: matches 293. .1 of consensus"
repeat_region 46958..47286
Query Match 13.1%; Score 18; DB 11; Length 89328;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 34505 CCACCGCACCTGGCTGAT 34522

RESULT 34
AC002126 89818 bp DNA PRI 15-AUG-1997
DEFINITION Homo sapiens DNA from chromosome 19-cosmids R30102:R29350:R27740
containing MEF2B, genomic sequence, complete sequence.
AC002126
AC002126.1 GI:2329908
VERSION
KEYWORDS HIG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89818)
Lamerdin,J.E.
Misc
Unpublished
2 (bases 1 to 89818)
Lamerdin,J.E., Kyle,A., Elliot,J., McCready,P.M., Kobayashi,A.,
Ow,D., Carrano,A.V. and Garcia,E.
Direct Submission
Submitted (15-AUG-1997) Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA 94550, USA
REMARK Genomic Structure of the Human MEF2B Gene
COMMENT Consensus sequence assembled from overlapping cosmids R30102 (bases
1- 36,529), R29350 (bases 29,365- 71,464), and R27740 (bases
54,941- 89,818).
FEATURES
Location/Qualifiers
source 1..89818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p12"
/chromosome="19"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/note="LL19NC03 library constructed at LLNL from
flow-sorted chromosomes from hybrid 5HL2-B, which carries
chromosome 19 as its only human chromosome"
complement(177..452)
/rpt_family="ALU"
800..1089
/rpt_family="ALU"
<1304..1611
/note="hypothetical partial orf most similar to orf from
C. elegans cosmid F43G9.3 (279755): putative exon, Xgrail
1.3c, excellent"
/codon_start=3
/product="R31240_1"
repeat_region
repeat_region
CDS
```

misc_feature	/protein_id="AAB86983.1" /db_xref="GI:2329910" /translation="YSGRRQYPFERMIFGACAGLIGOSASYPLDVVRRRMTAGVTG YPRASIAITLTIIVREEGAVKGLYKGSMMNVKGPVIAVGSFTTFDLNQILLRHLQS" 2171..2405 /note="BLASTN similarity to H67311 (1..236); match: 1, score: 4.1e-83; database searched: est; yu51b02.r1 Homo sapiens cDNA clone 229611 5'." complement(2403..2725) /note="BLASTN similarity to Z43933 (1..323); match: 0.98, score: 5.1e-132; database searched: est; H. sapiens partial cDNA sequence" 2404..2734 /note="BLASTN similarity to R11790 (1..332); match: 0.99, score: 2.8e-130; database searched: est; yf50e02.r1 Homo sapiens cDNA clone 25319 5'." 2683..2718 /note="BLASTN similarity to R10954 (3..38); match: 1, score: 1.5e-115; database searched: est; yf38f01.r1 Homo sapiens cDNA clone 129145 5'." complement(2705..2751) /note="BLASTN similarity to Z43933 (304..350); match: 0.82; score: 5.1e-132; database searched: est; H. sapiens partial cDNA sequence" 2706..2907 /note="BLASTN similarity to R10954 (25..226); match: 0.98, score: 1.5e-115; database searched: est; yf38f01.r1 Homo sapiens cDNA clone 129145 5'." 2734..2764 /note="BLASTN similarity to R11790 (333..363); match: 0.93; score: 2.8e-130; database searched: est; yf50e02.r1 Homo sapiens cDNA clone 25319 5'." 2744..2791 /note="BLASTN similarity to R11790 (344..391); match: 0.7, score: 2.3e-128; database searched: est; yf50e02.r1 Homo sapiens cDNA clone 25319 5'." 2829..3091 /note="BLASTN similarity to R22558 (1..263); match: 0.99, score: 2.2e-120; database searched: est; yh23h12.r1 Homo sapiens cDNA clone 130631 5'." 2830..3103 /note="BLASTN similarity to H03191 (1..274); match: 0.99, score: 3.8e-105; database searched: est; yj47b08.r1 Homo sapiens cDNA clone 151863 5'." 2913..3000 /note="BLASTN similarity to R10954 (234..321); match: 1, score: 1.5e-115; database searched: est; yf38f01.r1 Homo sapiens cDNA clone 129145 5'." complement(2923..2975) /note="BLASTN similarity to Z16004 (1..53); match: 0.94, score: 1.3e-114; database searched: est; H. sapiens partial cDNA sequence, clone 56B12" 2966..3276 /note="BLASTN similarity to R48116 (2..312); match: 0.98, score: 2.0e-134; database searched: est; yj63g05.r1 Homo sapiens cDNA clone 153464 5'." complement(2969..3179) /note="BLASTN similarity to Z16004 (48..258); match: 0.95, score: 1.3e-114; database searched: est; H. sapiens partial cDNA sequence, clone 56B12" 3088..3105 /note="BLASTN similarity to R22558 (261..278); match: 1, score: 2.2e-120; database searched: est; yh23h12.r1 Homo sapiens cDNA clone 130631 5'." complement(3165..3249) /note="BLASTN similarity to Z16004 (245..329); match: 0.92; score: 1.3e-114; database searched: est; H. sapiens partial cDNA sequence, clone 56B12" 3202..3233 /note="BLASTN similarity to R22558 (385..416); match: 1, score: 2.2e-120; database searched: est; yh23h12.r1 Homo sapiens cDNA clone 130631 5'." complement(3206..3245)	misc_feature /note="BLASTN similarity to H03990 (387..426); match: 0.9, score: 3.4e-142; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3243..3284) /note="BLASTN similarity to H03990 (347..388); match: 0.95; score: 3.4e-142; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3255..3284) /note="BLASTN similarity to R10902 (336..365); match: 0.76; score: 2.1e-117; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3266..3284) /note="BLASTN similarity to R22559 (356..374); match: 0.94; score: 1.1e-108; database searched: est; yh23h12.s1 Homo sapiens cDNA clone 130631 3'.">gb G30262 G30262 human STS SHGC-36640" complement(3270..3318) /note="BLASTN similarity to R10902 (301..349); match: 0.85; score: 2.8e-125; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3281..3318) /note="BLASTN similarity to H03990 (312..349); match: 0.73; score: 3.1e-97; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3297..3432) /note="BLASTN similarity to H03990 (197..332); match: 0.97; score: 3.4e-142; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3297..3321) /note="BLASTN similarity to R48002 (306..330); match: 0.92; score: 1.1e-95; database searched: est; yf63g05.s1 Homo sapiens cDNA clone 153464 3'." complement(3298..3621) /note="BLASTN similarity to Z39994 (1..324); match: 0.99, score: 6.2e-126; database searched: est; H. sapiens partial cDNA sequence" complement(3310..3362) /note="BLASTN similarity to R22559 (272..324); match: 0.83; score: 1.1e-108; database searched: est; yh23h12.s1 Homo sapiens cDNA clone 130631 3'.">gb G30262 G30262 human STS SHGC-36640" 3310..3339 /note="BLASTN similarity to R48116 (353..382); match: 0.8, score: 2.0e-134; database searched: est; yf63g05.r1 Homo sapiens cDNA clone 153464 5'." complement(3310..3432) /note="BLASTN similarity to R10902 (186..308); match: 0.97; score: 2.8e-125; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3322..3362) /note="BLASTN similarity to R48002 (262..302); match: 0.97; score: 1.1e-95; database searched: est; yf63g05.s1 Homo sapiens cDNA clone 153464 3'." 3332..3397 /note="BLASTN similarity to R48116 (379..444); match: 0.74; score: 2.0e-134; database searched: est; yf63g05.r1 Homo sapiens cDNA clone 153464 5'." complement(3336..3432) /note="BLASTN similarity to H68521 (192..288); match: 0.97; score: 2.9e-106; database searched: est; yu51b02.s1 Homo sapiens cDNA clone 229611 3'." complement(3339..3391) /note="BLASTN similarity to R22559 (242..294); match: 0.66; score: 5.6e-102; database searched: est; yh23h12.s1 Homo sapiens cDNA clone 130631 3'.">gb G30262 G30262 human STS SHGC-36640" complement(3371..3624) /note="BLASTN similarity to Z16003 (1..254); match: 0.95, score: 3.5e-89; database searched: est; H. sapiens partial cDNA sequence, clone 56B12" complement(3375..3629) /note="BLASTN similarity to R22559 (2..256); match: 0.98, score: 1.1e-108; database searched: est; yh23h12.s1 Homo
--------------	---	---

```
sapiens cDNA clone 130631 3'. >gb|G30262|G30262 human STS
SHGC-36640"
misc_feature
complement(3375..3620)
/Note="BLASTN similarity to R48002 (1..246); match: 0.91,
score: 1.1e-95; database searched: est; yj63g05.s1 Homo
sapiens cDNA clone 133464 3'."

Query Match      13.1%; Score 18; DB 11; Length 89818;
Best Local Similarity 100%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGAT 18
      |||
Db 8935 CCACCGCACCTGGCTGAT 8952

RESULT 35
AC009012/C
DEFINITION Homo sapiens chromosome 5 clone PI_360D11, *** SEQUENCING IN
PROGRESS ***, 60 unordered pieces.
ACCESSION AC009012
VERSION AC009012.1 GI:5686063
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 93841)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 93841)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 868: contig of 868 bp in length
* gap of unknown length
* 869 1905: contig of 1037 bp in length
* gap of unknown length
* 1906 2915: contig of 1010 bp in length
* gap of unknown length
* 2916 3805: contig of 890 bp in length
* gap of unknown length
* 3806 4552: contig of 747 bp in length
* gap of unknown length
* 4553 5406: contig of 854 bp in length
* gap of unknown length
* 5407 6278: contig of 872 bp in length
* gap of unknown length
* 6279 7171: contig of 893 bp in length
* gap of unknown length
* 7172 7847: contig of 676 bp in length
* gap of unknown length
* 7848 8794: contig of 947 bp in length
* gap of unknown length
* 8795 9778: contig of 984 bp in length
* gap of unknown length
* 9779 10920: contig of 1142 bp in length
* gap of unknown length
* 10921 11567: contig of 647 bp in length
* gap of unknown length
* 11568 12230: contig of 663 bp in length
* gap of unknown length

12231 13168: contig of 938 bp in length
gap of unknown length
13169 13825: contig of 657 bp in length
gap of unknown length
13826 14795: contig of 970 bp in length
gap of unknown length
14796 15595: contig of 800 bp in length
gap of unknown length
15596 16259: contig of 664 bp in length
gap of unknown length
16260 17239: contig of 980 bp in length
gap of unknown length
17240 18053: contig of 814 bp in length
gap of unknown length
18054 18859: contig of 806 bp in length
gap of unknown length
18860 19925: contig of 1066 bp in length
gap of unknown length
19926 20642: contig of 717 bp in length
gap of unknown length
20643 21396: contig of 754 bp in length
gap of unknown length
21397 22588: contig of 1192 bp in length
gap of unknown length
22589 23652: contig of 1064 bp in length
gap of unknown length
23653 25160: contig of 1508 bp in length
gap of unknown length
25161 26089: contig of 929 bp in length
gap of unknown length
26090 27249: contig of 1160 bp in length
gap of unknown length
27250 28390: contig of 1141 bp in length
gap of unknown length
28391 29660: contig of 1270 bp in length
gap of unknown length
29661 30582: contig of 922 bp in length
gap of unknown length
30583 31257: contig of 675 bp in length
gap of unknown length
31258 32197: contig of 940 bp in length
gap of unknown length
32198 33655: contig of 1458 bp in length
gap of unknown length
33656 34633: contig of 978 bp in length
gap of unknown length
34634 35188: contig of 555 bp in length
gap of unknown length
35189 36091: contig of 903 bp in length
gap of unknown length
36092 37188: contig of 1097 bp in length
gap of unknown length
37189 38179: contig of 991 bp in length
gap of unknown length
38180 39593: contig of 1414 bp in length
gap of unknown length
39594 41254: contig of 1861 bp in length
gap of unknown length
41255 42526: contig of 1272 bp in length
gap of unknown length
42527 43870: contig of 1344 bp in length
gap of unknown length
43871 45381: contig of 1511 bp in length
gap of unknown length
45382 46415: contig of 1034 bp in length
gap of unknown length
46416 47977: contig of 1562 bp in length
gap of unknown length
47978 49390: contig of 1413 bp in length
gap of unknown length
49391 51003: contig of 1613 bp in length
gap of unknown length
```

```
* 51004 52956: contig of 1953 bp in length
* 52957 gap of unknown length
* 54144: contig of 1188 bp in length
* 54145 gap of unknown length
* 55811: contig of 1667 bp in length
* 55812 gap of unknown length
* 56739: contig of 928 bp in length
* 56740 gap of unknown length
* 58330: contig of 1591 bp in length
* 58331 gap of unknown length
* 60299: contig of 1969 bp in length
* 62534: contig of 2235 bp in length
* 62535 gap of unknown length
* 64966: contig of 2432 bp in length
* 64967 gap of unknown length
* 68332: contig of 3366 bp in length
* 68333 93841: contig of 25509 bp in length.
FEATURES
SOURCE
1..93841
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="P1-360D11"
BASE COUNT 22734 a 23592 c 24325 g 23115 t 75 others
ORIGIN
Query Match 13.1%; Score 18; DB 41; Length 93841;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CTAAGTGAATCTTTCCT 76
|||||
Db 23119 CTAAGTGAATCTTTCCT 23102

RESULT 36
HSJ506G21/c 95425 bp DNA HTG 08-NOV-1996
LOCUS Homo sapiens chromosome X clone J506G21, *** SEQUENCING IN PROGRESS
DEFINITION
***, in unordered pieces.
ACCESSION 282213
VERSION 282213.1 GI:1666405
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 95425)
Direct Submission
Submitted (06-NOV-1996) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
Order of segments is not known; 800 n/s separate segments.
Unfinished sequence: dj506G2 Contig_ID: 01828 Length: 797 bp
Unfinished sequence: dj506G2 Contig_ID: 01922 Length: 37027 bp
Unfinished sequence: dj506G2 Contig_ID: 01957 Length: 1258 bp
Unfinished sequence: dj506G2 Contig_ID: 00779 Length: 53943 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Location/Qualifiers
source
1..95425
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="J506G21"
BASE COUNT 30849 a 19403 c 17797 g 24967 t 2409 others
ORIGIN
```

```
Query Match 13.1%; Score 18; DB 32; Length 95425;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGAT 18
|||||
Db 5906 CCACCGCACCTGGCTGAT 5889

RESULT 37
AC011336 96099 bp DNA HTG 06-OCT-1999
LOCUS Homo sapiens chromosome 5 clone CIT-HSPC_250I13, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
ACCESSION AC011336
VERSION AC011336.1 GI:6013604
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 96099)
Direct Submission
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 705: contig of 705 bp in length
gap of unknown length
706 1367: contig of 662 bp in length
gap of unknown length
1368 2143: contig of 776 bp in length
gap of unknown length
2144 2940: contig of 797 bp in length
gap of unknown length
2941 4813: contig of 1873 bp in length
gap of unknown length
4814 7980: contig of 3167 bp in length
gap of unknown length
7981 49500: contig of 41520 bp in length
gap of unknown length
49501 96099: contig of 46599 bp in length.
Location/Qualifiers
source
1..96099
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CIT-HSPC_250I13"
BASE COUNT 23213 a 24003 c 24571 g 24303 t 9 others
ORIGIN

Query Match 13.1%; Score 18; DB 41; Length 96099;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 CTAAGTGAATCTTTCCT 76
|||||
Db 68149 CTAAGTGAATCTTTCCT 68166
```

RESULT 38

AC013370 100091 bp DNA HTG 20-NOV-1999
 LOCUS Homo sapiens clone RP11-12D16, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC013370
 ACCESSION AC013370.3 GI:6456269
 VERSION
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 100091)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-12D16

REFERENCE

2 (bases 1 to 100091)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
 Galagan,J., Gardynda,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission

TITLE

JOURNAL

Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 20, 1999 this sequence version replaced gi:6425738.

COMMENT

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Project Information

Center project name: L3253

Center clone name: 12_D_16

 * NOTE: This record contains 110 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be generic and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 910: contig of 910 bp in length

gap of unknown length

911 1846: contig of 936 bp in length

gap of unknown length

1847 2768: contig of 922 bp in length

gap of unknown length

2769 3751: contig of 983 bp in length

gap of unknown length

3752 4608: contig of 857 bp in length

gap of unknown length

4609 5507: contig of 899 bp in length

gap of unknown length

5508 6398: contig of 891 bp in length

gap of unknown length

6399 7310: contig of 912 bp in length

gap of unknown length

7311 8239: contig of 929 bp in length

gap of unknown length

8240 9173: contig of 934 bp in length

gap of unknown length

9174 10107: contig of 934 bp in length
 gap of unknown length
 10108 11014: contig of 907 bp in length
 gap of unknown length
 11015 11934: contig of 920 bp in length
 gap of unknown length
 11935 12855: contig of 921 bp in length
 gap of unknown length
 12856 13746: contig of 891 bp in length
 gap of unknown length
 13747 14742: contig of 996 bp in length
 gap of unknown length
 14743 15668: contig of 926 bp in length
 gap of unknown length
 15669 16623: contig of 955 bp in length
 gap of unknown length
 16624 17547: contig of 924 bp in length
 gap of unknown length
 17548 18486: contig of 939 bp in length
 gap of unknown length
 18487 19433: contig of 947 bp in length
 gap of unknown length
 19434 20378: contig of 945 bp in length
 gap of unknown length
 20379 21271: contig of 893 bp in length
 gap of unknown length
 21272 22176: contig of 905 bp in length
 gap of unknown length
 22177 23090: contig of 914 bp in length
 gap of unknown length
 23091 24002: contig of 912 bp in length
 gap of unknown length
 24003 24930: contig of 928 bp in length
 gap of unknown length
 24931 25870: contig of 940 bp in length
 gap of unknown length
 25871 26812: contig of 942 bp in length
 gap of unknown length
 26813 27775: contig of 963 bp in length
 gap of unknown length
 27776 28688: contig of 913 bp in length
 gap of unknown length
 28689 29669: contig of 981 bp in length
 gap of unknown length
 29670 30581: contig of 912 bp in length
 gap of unknown length
 30582 31525: contig of 944 bp in length
 gap of unknown length
 31526 32441: contig of 916 bp in length
 gap of unknown length
 32442 33365: contig of 924 bp in length
 gap of unknown length
 33366 34290: contig of 925 bp in length
 gap of unknown length
 34291 35205: contig of 915 bp in length
 gap of unknown length
 35206 36136: contig of 931 bp in length
 gap of unknown length
 36137 37079: contig of 943 bp in length
 gap of unknown length
 37080 38018: contig of 939 bp in length
 gap of unknown length
 38019 38963: contig of 945 bp in length
 gap of unknown length
 38964 39855: contig of 892 bp in length
 gap of unknown length
 39856 40787: contig of 932 bp in length
 gap of unknown length
 40788 41729: contig of 942 bp in length
 gap of unknown length
 41730 42618: contig of 889 bp in length
 gap of unknown length

REFERENCE AUTHORS TITLE JOURNAL

1 (bases 1 to 109210)

Blakey, S.

Direct Submission

Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, UK. E-mail enquiries: humquery@sanger.ac.uk

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 18, 1999 this sequence version replaced gi:5708178.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence is

the entire insert of clone 863C7. This sequence has been finished

according to sequence map criteria as follows. An attempt is made

to resolve all sequencing problems, such as compressions and

repeats, but not necessarily within known annotated human repeat

sequence elements (e.g. Alu). Where the sequence is ambiguous,

there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome 20, constructed by the Sanger Centre Chromosome 20

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

863C7 is from the library RPCI-5 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2.

FEATURES

source

1..109210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="p12.3-13"
/clone_lib="RPCI-5"
/clone="RP5-863C7"

BASE COUNT 32132 a 23899 c 22881 g 30298 t

ORIGIN

Query Match 13.1%; Score 18; DB 11; Length 109210;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 99544 CCACCGCACCTGGCTGAT 99527

RESULT 41

HS112K5/c

LOCUS

DEFINITION

Human DNA sequence from clone 112K5 on chromosome Xp11, complete

sequence.

ACCESSION 285987

VERSION 285987.13

KEYWORDS GI:5531490

HTG.

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 109891)

Pearce, A.

Direct Submission

Submitted (19-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jul 19, 1999 this sequence version replaced gi:5459255.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL

This sequence is the entire insert of clone 112K5. This sequence

has been finished according to sequence map criteria as follows. An

attempt is made to resolve all sequencing problems, such as

compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome X, constructed by the Sanger Centre Chromosome X

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

112K5 is from the library RPCI1 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2.

FEATURES

source

1..109891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11"
/clone="RP1-112K5"
/clone_lib="RPCI-1"

BASE COUNT 29357 a 25957 c 26406 g 28171 t

ORIGIN

Query Match 13.1%; Score 18; DB 10; Length 109891;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 78273 CCACCGCACCTGGCTGAT 78256

RESULT 42

AC010679_2

WPCOMMENT

Sequence split into 4 fragments LOCUS AC010679 Accession AC010679

Fragment Name Begin End

AC010679_0 1 110000

AC010679_1 100001 210000

AC010679_2 200001 310000

AC010679_3 300001 350997

Continuation (3 of 4) of AC010679 from base 200001 (AC010679 Homo sapiens clone NH012

Query Match

Best Local Similarity

Matches

13.1%; Score 18; DB 41; Length 110000;

100.0%; Pred. No. 5.8;

18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 78298 CCACCGCACCTGGCTGAT 78315

RESULT 43

AC015845_0/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC015845 Accession AC015845

Fragment Name Begin End

AC015845_0 1 110000

AC015845_1 100001 210000

AC015845_2 200001 310000

AC015845_3 300001 410000

AC015845_4 400001 510000

```

AC015845.5      500001      573550
LOCUS      AC015845      573550 bp      DNA      HTG      17-NOV-1999
DEFINITION      Homo sapiens chromosome 17 clone RP11-343K8 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION      AC015845
VERSION      AC015845.1      GI:6446823
KEYWORDS      HTG: HTGS_PSEO; NULL.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 573550)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Homo sapiens chromosome 17, clone RP11-343K8
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 573550)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Balwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckghalter, B.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collamore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
TITLE      Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
JOURNAL      Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      All repeats were identified using RepeatMasker:
      Smit, A.F.A. & Green, P. (1996-1997)
      http://ftp.genome.washington.edu/RN/RepeatMasker.html
      ----- Project Information
      Center project name: L479
      Center clone name: 343_K_8
      -----
      * NOTE: This record contains 687 individual
      * sequencing reads that have not been assembled into
      * contigs. Runs of N are used to separate the reads
      * and the order in which they appear is completely
      * arbitrary. Low-pass sequence sampling is useful for
      * identifying clones that may be gene-rich and allows
      * overlap relationships among clones to be deduced.
      * However, it should not be assumed that this clone
      * will be sequenced to completion. In the event that
      * the record is updated, the accession number will
      * be preserved.
      *
      * 1      780:      contig of 780 bp in length
      *          gap of unknown length
      *
      * 781      1566:      contig of 786 bp in length
      *
      * 1567      2334:      contig of 768 bp in length
      *
      * 2335      3108:      contig of 774 bp in length
      *
      * 3109      3900:      contig of 792 bp in length
      *
      * 3901      4700:      contig of 800 bp in length
      *
      * 4701      5482:      contig of 782 bp in length
      *
      * 5483      6257:      contig of 775 bp in length
      *
      * 6258      7034:      contig of 777 bp in length
      *
      * 7035      7795:      contig of 761 bp in length
      *
      * 7796      8572:      contig of 777 bp in length
      *          gap of unknown length
      *
      * 9355:      contig of 783 bp in length
      *          gap of unknown length
      *
      * 10127:      contig of 772 bp in length
      *          gap of unknown length
      *
      * 10892:      contig of 765 bp in length
      *          gap of unknown length
      *
      * 11643:      contig of 751 bp in length
      *          gap of unknown length
      *
      * 12435:      contig of 792 bp in length
      *          gap of unknown length
      *
      * 13236:      contig of 801 bp in length
      *          gap of unknown length
      *
      * 14022:      contig of 786 bp in length
      *          gap of unknown length
      *
      * 14808:      contig of 786 bp in length
      *          gap of unknown length
      *
      * 15592:      contig of 784 bp in length
      *          gap of unknown length
      *
      * 16375:      contig of 783 bp in length
      *          gap of unknown length
      *
      * 17118:      contig of 743 bp in length
      *          gap of unknown length
      *
      * 17909:      contig of 791 bp in length
      *          gap of unknown length
      *
      * 18698:      contig of 789 bp in length
      *          gap of unknown length
      *
      * 19488:      contig of 790 bp in length
      *          gap of unknown length
      *
      * 20168:      contig of 680 bp in length
      *          gap of unknown length
      *
      * 20951:      contig of 783 bp in length
      *          gap of unknown length
      *
      * 21731:      contig of 780 bp in length
      *          gap of unknown length
      *
      * 22519:      contig of 788 bp in length
      *          gap of unknown length
      *
      * 23304:      contig of 785 bp in length
      *          gap of unknown length
      *
      * 24077:      contig of 773 bp in length
      *          gap of unknown length
      *
      * 24868:      contig of 791 bp in length
      *          gap of unknown length
      *
      * 25668:      contig of 800 bp in length
      *          gap of unknown length
      *
      * 26451:      contig of 783 bp in length
      *          gap of unknown length
      *
      * 27234:      contig of 783 bp in length
      *          gap of unknown length
      *
      * 27992:      contig of 758 bp in length
      *          gap of unknown length
      *
      * 28776:      contig of 784 bp in length
      *          gap of unknown length
      *
      * 29566:      contig of 790 bp in length
      *          gap of unknown length
      *
      * 30372:      contig of 806 bp in length
      *          gap of unknown length
      *
      * 31172:      contig of 800 bp in length
      *          gap of unknown length
      *
      * 31959:      contig of 787 bp in length
      *          gap of unknown length
      *
      * 32743:      contig of 784 bp in length
      *          gap of unknown length
      *
      * 33545:      contig of 802 bp in length
      *          gap of unknown length
      *
      * 34329:      contig of 784 bp in length
      *          gap of unknown length
      *
      * 35061:      contig of 732 bp in length
      *          gap of unknown length
      *
      * 35845:      contig of 784 bp in length
      *          gap of unknown length
      *
      * 36632:      contig of 787 bp in length
      *          gap of unknown length
      *
      * 37433:      contig of 801 bp in length
  
```

* 37434 38226: contig of 793 bp in length
* 38227 38965: contig of 739 bp in length
* 38966 39753: contig of 788 bp in length
* 39754 40522: contig of 769 bp in length
* 40523 41305: contig of 783 bp in length
* 41306 42089: contig of 784 bp in length
* 42090 42885: contig of 796 bp in length
* 42886 43653: contig of 768 bp in length
* 43654 44394: contig of 741 bp in length
* 44395 45093: contig of 699 bp in length
* 45094 45881: contig of 788 bp in length
* 45882 46663: contig of 782 bp in length
* 46664 47449: contig of 786 bp in length
* 47450 48232: contig of 783 bp in length
* 48233 49013: contig of 781 bp in length
* 49014 49795: contig of 782 bp in length
* 49796 50575: contig of 780 bp in length
* 50576 51383: contig of 808 bp in length
* 51384 52240: contig of 857 bp in length
* 52241 53047: contig of 807 bp in length
* 53048 53847: contig of 800 bp in length
* 53848 54624: contig of 777 bp in length
* 54625 55406: contig of 782 bp in length

Query Match 13.1%; Score 18; DB 43; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCCTGGCTGAT 18
|||||

Db 85101 CCACCGCACCCTGGCTGAT 85084

RESULT 44
AC015845_2/c
WPCOMMENT

Sequence split into 6 fragments LOCUS AC015845 Accession AC015845

Fragment Name	Begin	End
AC015845_0	1	110000
AC015845_1	100001	210000
AC015845_2	200001	310000
AC015845_3	300001	410000
AC015845_4	400001	510000
AC015845_5	500001	573550

Continuation (3 of 6) of AC015845 from base 200001 (AC015845 Homo sapiens chromosome 17

Query Match 13.1%; Score 18; DB 43; Length 110000;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCCTGGCTGAT 18
|||||

Db 29319 CCACCGCACCCTGGCTGAT 29302

RESULT 45

AC013673 121345 bp DNA HTG 13-NOV-1999
LOCUS Homo sapiens clone 20_I_21, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC013673
ACCESSION AC013673.1 GI:6403568
VERSION
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 121345)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens, clone 20_I_21

JOURNAL

Unpublished

REFERENCE

AUTHORS

2 (bases 1 to 121345)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguski, R., Boughey, L., Boughey, L.,
Brown, A., Castle, A., Collins, S., Collins, S., Collins, S., Collins, S.,
Cooke, P., DeArnell, K., Dewar, K., Domino, M., Donnelly, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Gallagher, J., Gardyna, S., Grant, G., Hagan, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 153 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	824:	contig of 824 bp in length
*		gap of unknown length
*	825	1622: contig of 798 bp in length
*		gap of unknown length
*	1623	2400: contig of 778 bp in length
*		gap of unknown length
*	2401	3201: contig of 801 bp in length
*		gap of unknown length
*	3202	3986: contig of 785 bp in length
*		gap of unknown length
*	3987	4779: contig of 793 bp in length
*		gap of unknown length
*	4780	5590: contig of 811 bp in length
*		gap of unknown length
*	5591	6387: contig of 797 bp in length
*		gap of unknown length
*	6388	7199: contig of 812 bp in length
*		gap of unknown length
*	7200	7948: contig of 749 bp in length
*		gap of unknown length

```

* 7949 8737: contig of 789 bp in length
* gap of unknown length
* 8738 9543: contig of 806 bp in length
* gap of unknown length
* 9544 10346: contig of 803 bp in length
* gap of unknown length
* 10347 11190: contig of 844 bp in length
* gap of unknown length
* 11191 12006: contig of 816 bp in length
* gap of unknown length
* 12007 12797: contig of 791 bp in length
* gap of unknown length
* 12798 13571: contig of 774 bp in length
* gap of unknown length
* 13572 14425: contig of 854 bp in length
* gap of unknown length
* 14426 15224: contig of 799 bp in length
* gap of unknown length
* 15225 16013: contig of 789 bp in length
* gap of unknown length
* 16014 16805: contig of 792 bp in length
* gap of unknown length
* 16806 17621: contig of 816 bp in length
* gap of unknown length
* 17622 18451: contig of 830 bp in length
* gap of unknown length
* 18452 19254: contig of 803 bp in length
* gap of unknown length
* 19255 20056: contig of 802 bp in length
* gap of unknown length
* 20057 20850: contig of 794 bp in length
* gap of unknown length
* 20851 21634: contig of 784 bp in length
* gap of unknown length
* 21635 22419: contig of 785 bp in length
* gap of unknown length
* 22420 23210: contig of 791 bp in length
* gap of unknown length
* 23211 23993: contig of 783 bp in length
* gap of unknown length
* 23994 24786: contig of 793 bp in length
* gap of unknown length
* 24787 25578: contig of 792 bp in length
* gap of unknown length
* 25579 26393: contig of 815 bp in length
* gap of unknown length
* 26394 27192: contig of 799 bp in length
* gap of unknown length
* 27193 28002: contig of 810 bp in length
* gap of unknown length
* 28003 28832: contig of 830 bp in length
* gap of unknown length
* 28833 29638: contig of 806 bp in length
* gap of unknown length
* 29639 30411: contig of 773 bp in length
* gap of unknown length
* 30412 31191: contig of 780 bp in length
* gap of unknown length
* 31192 32000: contig of 809 bp in length
* gap of unknown length
* 32001 32780: contig of 780 bp in length
* gap of unknown length
* 32781 33600: contig of 820 bp in length
* gap of unknown length
* 33601 34397: contig of 797 bp in length
* gap of unknown length
* 34398 35193: contig of 796 bp in length
* gap of unknown length
* 35194 35976: contig of 783 bp in length
* gap of unknown length
* 35977 36772: contig of 796 bp in length
* gap of unknown length
* 36773 37564: contig of 792 bp in length

```

```

* 37565 38346: contig of 782 bp in length
* gap of unknown length
* 38347 39123: contig of 777 bp in length
* gap of unknown length
* 39124 39921: contig of 798 bp in length
* gap of unknown length
* 39922 40723: contig of 802 bp in length
* gap of unknown length
* 40724 41523: contig of 800 bp in length
* gap of unknown length
* 41524 42358: contig of 835 bp in length
* gap of unknown length
* 42359 43174: contig of 816 bp in length
* gap of unknown length
* 43175 43989: contig of 815 bp in length
* gap of unknown length
* 43990 44779: contig of 790 bp in length
* gap of unknown length
* 44780 45569: contig of 790 bp in length
* gap of unknown length
* 45570 46371: contig of 802 bp in length
* gap of unknown length
* 46372 47169: contig of 798 bp in length
* gap of unknown length
* 47170 47957: contig of 788 bp in length
* gap of unknown length
* 47958 48772: contig of 815 bp in length
* gap of unknown length
* 48773 49548: contig of 776 bp in length
* gap of unknown length
* 49549 50338: contig of 790 bp in length
* gap of unknown length
* 50339 51143: contig of 805 bp in length
* gap of unknown length
* 51144 51943: contig of 800 bp in length
* gap of unknown length
* 51944 52666: contig of 723 bp in length
* gap of unknown length
* 52667 53437: contig of 771 bp in length
* gap of unknown length
* 53438 54240: contig of 803 bp in length
* gap of unknown length
* 54241 55042: contig of 802 bp in length
* gap of unknown length
* 55043 55827: contig of 785 bp in length
* gap of unknown length
* 55828 56632: contig of 805 bp in length
* gap of unknown length
* 56633 57449: contig of 817 bp in length
* gap of unknown length
* 57450 58278: contig of 829 bp in length
* gap of unknown length
* 58279 59053: contig of 775 bp in length
* gap of unknown length
* 59054 59838: contig of 785 bp in length
* gap of unknown length
* 59839 60636: contig of 798 bp in length
* gap of unknown length
* 60637 61415: contig of 779 bp in length
* gap of unknown length
* 61416 62222: contig of 807 bp in length

```

Query Match 13.1% Score 18; DB 43; Length 121345;
 Best Local Similarity 100.0%; Pred. No. 5.8;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 CTTCCAAGCTAAACAAT 100
 |||||
 DB 10859 CTTCCAAGCTAAACAAT 10876

Search completed: April 6, 2000, 21:20:42
Job time: 50627 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2000, 11:35:47 ; Search time 6420.76 Seconds

(without alignments)
161.123 Million cell updates/sec

Title: US-09-090-672B-12

Perfect score: 274

Sequence: 1 CGTTACAGATCTCTTCG.....GACTATCAAAATGAAAAAT 274

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		% Match	Length	Score			
1	198.6	72.5	405	21	T98890		T98890 ye65f09.r1
2	138	50.4	493	23	H18435		H18435 ym43b05.s1

c 3 127.4 46.5 302 20 F03205
4 117.8 43.0 390 39 AA878923
5 113 41.2 318 30 AA197894
6 113 41.2 459 39 AA873942
7 111.4 40.7 731 74 AA211128
8 88.4 32.3 391 33 AA432984
9 81.8 29.9 387 22 H09173
10 74.6 27.2 274 32 AA334229
11 64.6 23.6 559 34 AA488906
12 64.2 23.4 259 20 T10350
13 63.6 23.2 596 104 AQ587717
14 62.8 22.9 449 43 A1169903
15 53.8 19.6 614 79 FR0008178
16 47.6 17.4 1203 80 CNS015WU
17 45.4 16.6 354 22 R45421
18 41.8 15.3 1201 80 CNS0167M
19 41.6 15.2 624 105 AQ620120
20 41.2 15.0 242 45 A1394828
21 41 15.0 483 49 A1632118
22 40.8 14.9 583 104 AQ570963
23 40.6 14.8 437 38 AA755877
24 40.4 14.7 415 35 AA550066
25 40 14.6 423 88 AQ836402
26 40 14.6 488 91 AQ129266
27 39.8 14.5 1101 80 CNS016DT
28 39.6 14.5 1101 79 CNS008X3
29 39.4 14.4 467 80 CNS018G8
30 39.4 14.4 1101 80 CNS0176D
31 39 14.2 1101 80 CNS0142R
32 39 14.2 1204 80 CNS016E2
33 38.8 14.2 346 60 A1783824
34 38.4 14.0 561 91 AQ139258
35 38.4 14.0 618 100 AQ326393
36 38.4 14.0 654 99 AQ259988
37 38.4 14.0 928 79 CNS00DKY
38 38.4 14.0 1101 79 CNS003B6
39 37.8 13.8 583 45 A1394910
40 37.6 13.7 259 73 AV373585
41 37.4 13.6 455 45 A1395037
42 37.2 13.6 229 45 A1394956
43 37.2 13.6 778 79 CNS010L9
44 37 13.5 299 61 A1860086
45 37 13.5 441 61 A1811918

ALIGNMENTS

RESULT 1
T98890 405 bp mRNA EST 31-MAR-1995
LOCUS ye65f09.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:122633 5', mRNA sequence.
ACCESSION T98890
VERSION T98890.1 GI:748627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfling, T., Soares, M., Tan, F.,
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Insert Size: 1065
High quality sequence stops: 279 Source: IMAGE Consortium, LNL This
clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert length: 1065 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 279.
Location/Qualifiers
1. 405
/organism="Homo sapiens"
/db_xref="GDB:475178"
/db_xref="taxon:9606"
/clone="IMAGE:122633"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/note="Organ: Liver and Spleen; Vector: pT7r3D (Pharmacia)
with a modified polylinker; Site:1: Pac I; Site:2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAAGAATAATAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7r3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 116 a 68 c 133 t 4 others
ORIGIN
Query Match 72.5%; Score 198.6; DB 21; Length 405;
Best Local Similarity 98.7%; Pred. No. 2.1e-36;
Matches 220; Conservative 1; Mismatches 0; Indels 2; Gaps 2;
QY 31 GAATCAAAAGGGATCGCTATATACACATACCAACTGATAAATAGTAGATTCT 90
Db 2 GAATCAAAAGGGATCGCTATATACACATACCAACTGATAAATAGTAGATTCT 61
QY 91 GTGTYCTGCTTATAGACCATGTTTGTAGTAGTAAGAGAGAAACTTCCTATATCTGAA 150
Db 62 GTGTCCTGCTTATAGACCATGTTTGTAGTAGTAAGAGAGAAACTTCCTATATCTGAA 121
QY 151 ACAGCTTAACATTTTACAAAATTTAGTTTCTTTTGTAGAGTCTTATCTGTA-GCTAT 209
Db 122 ACAGCTTAACATTTTACAAAATTTAGTTTCTTTTGTAGAGTCTTATCTGTA-GCTAT 181
QY 210 ATACAGTTCATGCTGATTGA-GCATTTGTTTCAGGAGTAAAG 251
Db 182 ATACAGTTCATGCTGATTGAGCATTTGTTTCAGGAGTAAAG 224
RESULT 2
LOCUS H18435/c
DEFINITION ym43b05.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:50988 3', mRNA sequence.
ACCESSION H18435
VERSION H18435.1 GI:884675
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfling, T., Soares, M., Tan, F.,
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 18, 1995 this sequence version replaced gi:811454.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1071
High quality sequence steps: 327
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1071 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 327.

FEATURES

Location/Qualifiers
1..493
/organism="Homo sapiens"
/db_xref="GDB:423798"
/db_xref="taxon:9606"
/clone="IMAGE:50988"
/clone_lib="Soares infant brain L1NB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site.1: Not
I; Site.2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
ACTGGAGAAATCGGCGCGGAGGAATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
146 a 86 c 106 g 148 t 7 others

BASE COUNT
ORIGIN

Query Match 50.4%; Score 138; DB 23; Length 493;
Best Local Similarity 77.3%; Pred. No. 1.7e-22;
Matches 218; Conservative 0; Mismatches 54; Indels 10; Gaps 4;

QY 3 TTATACGATTCTTTCGGCGTGGGAACTACAAAGGATCGGTG-CCTATATACACA 61
|||||
Db 466 TTAACNAACTCCCGCGGAGACGGGAACTACAAAGGACGGTGCCTATATACACA 407
QY 62 ATACCAACTGTATATATCTAGATCTGCTGYTCTGTCTATAGACCATGTTGT --- 118
|||||
Db 406 ATACCAACTGTATATATACACCTAGATCTGCTGTGCTCCGCCCTAATAGACCATGTTGT 347
QY 118 -AGTAGTAAGAGAAAACCTTC--TATATCTGAAACAGCCCTAACATTTTACAAAAT 173
|||||
Db 346 AAGTAGTAGAGGAAACTTCCTATATCTTGAACAGCCCTAACATTTTACAAAAT 287
QY 174 TTAGTTTCTTTTATAGAGCTTATCCTGTAG-CTATATAACAGTTCATGCTCGATTAG 232
|||||
Db 286 TTAGTTTCTTTTATAGAGCTTATCCTGTAGCCTATATAACAGTTCATGCTCGATTAG 227
QY 233 CATTTGTTACGAGTAAGCTGGAACATATGAAAATGAAAAT 274
|||||
Db 226 CATTTGTTACGAGTAAGCTGGAACATATGAAAATGAAAAT 185

RESULT 3

LOCUS F03205 302 bp mRNA EST 02-FEB-1995
DEFINITION HSC1PF072 normalized infant brain cDNA Homo sapiens cDNA clone
c-lpf07 3', mRNA sequence.
ACCESSION F03205
VERSION F03205.1 GI:646762
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

AUTHORS

Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 302)
Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL

MEDLINE

COMMENT

95277534

Contact: Genethon

Genexpress-genethon

Genethon Centre de recherche sur le Genome Humain

1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpress@genethon.fr

Single read, removed at sequence 5'end

Genexpress_library_idt: C; Genexpress_sequence_idt: alc-lpf07

Seq primer: (-21)M13.universal.

FEATURES

Source

1..302

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="c-lpf07"

/clone_lib="normalized infant brain cDNA"

/sex="Female"

/tissue_type="total brain"

/dev_stage="3 months old"

/note="Organ: brain; Vector: Lfamid BA; Site.1: HindIII;
Site.2: NotI; sex:Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
Lfamid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

111 a 54 c 52 g 84 t 1 others

BASE COUNT
ORIGIN

Query Match 46.5%; Score 127.4; DB 20; Length 302;
Best Local Similarity 99.2%; Pred. No. 5.2e-20;
Matches 128; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 146 CTGAACAGCCTACATTTACAAATTTAGTTTCTTTTAGAGTCTATCCTGTAG 205
|||||
Db 302 CTGAACAGCCTACATTTACAAATTTAGTTTCTTTTAGAGTCTATCCTGTAG 243
QY 206 CTATATACAGTTCATGCTGATTAGCATTTGTTACAGTAAGCTGGAACATGAAA 265
|||||
Db 242 CTATATACAGTTCATGCTGATTAGCATTTGTTACAGTAAGCTGGAACATGAAA 183
QY 266 ATTGAAAAT 274
|||||
Db 182 ATTGAAAAT 174

RESULT 4

LOCUS AA878923

DEFINITION oJ25d11.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1493205 3',
similar to TR:O14432 O14432 TUPI.1, mRNA sequence.

ACCESSION AA878923

VERSION AA878923.1 GI:2987888

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 390)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index
 Unpublished (1997)
 On Jan 19, 1998 this sequence version replaced gi:2152443.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.lnl.gov/bbrp/image/image.html

Insert Length: 1846 Std Error: 0.00
 Seq primer: -40ml3 fwd ET from Amersham
 High quality sequence stop: 365.

FEATURES

source
 Location/Qualifiers
 1..390
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1493205"
 /clone_lib="NCI-CGAP_Kid5"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGGAAGATTCGGCGCGCATATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 110 a 86 c 87 g 107 t
 ORIGIN

Query Match 43.0%; Score 117.8; DB 39; Length 390;
 Best Local Similarity 96.7%; Pred. No. 8.1e-18;
 Matches 119; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 CGTTTACAGATTCCTTCGGCGTGGGATGGAACCTACAAAGGATCGTGCTATATAC 60
 Db 30 CGTTTACAGATTCCTTCGGCGAGCGGTGGAACCTACAAAGGATCGTGCTATATAC 89
 QY 61 AATACCAAACTTGATAAATCTAGATTCGTGTCTGTATAGACCATGTTGTAGT 120
 Db 90 AATACCAAACTTGATAAATCTAGATTCGTGTCTGTATAGACCATGTTGTAGT 149
 QY 121 AGG 123
 Db 150 AGG 152

RESULT 5
 LOCUS AA197894 318 bp mRNA EST 12-MAR-1997
 DEFINITION mv05e02.r1 GuayWoodford Beier mouse kidney day 0 Mus musculus cDNA clone IMAGE:654170 5', mRNA sequence.

ACCESSION AA197894
 VERSION AA197894.1 GI:1793528
 KEYWORDS EST.

SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 318)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

TITLE
 JOURNAL
 COMMENT

Waterston, R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1394222.
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:400018

Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 297.

FEATURES

source
 Location/Qualifiers
 1..318
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:654170"
 /clone_lib="GuayWoodford Beier mouse kidney day 0"
 /tissue_type="kidney"
 /dev_stage="newborn (day 0)"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: kidney; Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' Library provided Lisa Guay-Woodford."

BASE COUNT 89 a 74 c 73 g 82 t
 ORIGIN

Query Match 41.2%; Score 113; DB 30; Length 318;
 Best Local Similarity 94.3%; Pred. No. 1.1e-16;
 Matches 116; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGTTTACAGATTCCTTCGGCGTGGGATGGAACCTACAAAGGATCGTGCTATATAC 60
 Db 112 CGTTTACAGATTCCTTCGGCGAGCGGTGGAACCTACAAAGGATCGTGCTATATAC 171
 QY 61 AATACCAAACTTGATAAATCTAGATTCGTGTCTGTATAGACCATGTTGTAGT 120
 Db 172 AATACCAAACTTGATAAATCTAGATTCGTGTCTGTATAGATCATGTTGTAGT 231
 QY 121 AGG 123
 Db 232 AGG 234

RESULT 6
 LOCUS AA873942 459 bp mRNA EST 19-MAR-1998

DEFINITION vw86a03.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone IMAGE:1261804 5' similar to TR:014432 O14432 TUPL.1; mRNA sequence.

ACCESSION AA873942
 VERSION AA873942.1 GI:2978631
 KEYWORDS EST.

SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 459)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 The WashU-HHMI Mouse EST Project

JOURNAL
COMMENT

Unpublished (1996)
On May 18, 1995 this sequence version replaced gi:811146.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:664336
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 453.

FEATURES
source

1. 459
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1261804"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor
sequence: 5' CTCGATTTTTTTTTTTTT 3"
131 a 95 c 102 g 131 t

BASE COUNT
ORIGIN

Query Match 41.2%; Score 113; DB 39; Length 459;
Best Local Similarity 94.3%; Pred. No. 1e-16;
Matches 116; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CGTTTACAGATCTCTTCGGCGTGGGAACATACAAAGGGATCGGCTATATCAC 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 CGTTTACAGATCTCTCTCGCGAGCGGTGGAACATACAAAGGGATCGGCTATATCAC 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 AATACCAACTTGATAATATCTAGATTCGTGTCTGTCTTATAGACCATGTTTGAGT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 AATACCAACTTGACAATTAATCTAGATTCGTGTCTGTCTTATAGATCATGTTTGAGT 160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 121 AGG 123
|||
Db 161 AGG 163

RESULT 7

AW211128 731 bp mRNA EST 03-DEC-1999
LOCUS ul24h09.v1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IMAGE:209297 5' similar to TR:O31259 O31259 VANADIUM
CHLOROPEROXIDASE. ; mRNA sequence.

AW211128
VERSION AW211128.1 GI:6517068
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 731)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,I., Jackson,I., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)

COMMENT

On Mar 10, 1998 this sequence version replaced gi:2948074.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:996229

Seq primer: custom primer used
High quality sequence stop: 521.

FEATURES
source

1. 731
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:209297"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(df) primer
[ATGTGGCTTTTITTTTTTTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGACA."
197 a 165 c 180 g 189 t

BASE COUNT
ORIGIN

Query Match 40.7%; Score 111.4; DB 74; Length 731;
Best Local Similarity 93.5%; Pred. No. 2.2e-16;
Matches 115; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CGTTTACAGATTCCTTTCGGCGTGGGAACATACAAAGGGATCGGCTATATCAC 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 CGTTTACAGATCTCTCTCGCGAGCGGTGGAACATACAAAGGGATCGGCTATATCAC 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 AATACCAAACTTGATAATATCTAGATTCGTGTCTGTCTTATAGACCATGTTTGAGT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 AATACCAAACTTGACAATAATCTAGATTCGTGTCTGTCTTACCTATAGATCATGTTTGAGT 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 121 AGG 123
|||
Db 465 AGG 467

RESULT 8

AA432984 391 bp mRNA EST 04-AUG-1997
LOCUS ve87b12.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
DEFINITION IMAGE:833183 5', mRNA sequence.

AA432984
ACCESSION AA432984
VERSION AA432984.1 GI:2116414
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 391)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geiseli,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

TITLE
JOURNAL
COMMENT

Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
On May 9, 1995 this sequence version replaced gi:804216.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:493399

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 358.

FEATURES

source
Location/Qualifiers
1..391
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:833183"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH108"

/note="Organ: embryo; Vector: pBluescribe (modified);
Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
mRNA prepared from 13 500 2-cell stage embryos. Primer:
SalI(4T): 5'-CGTGCAGCGTCGACGCTTTTCTTTT-3'. CDNAS
were cloned into the MluI/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."

BASE COUNT 118 a 83 c 96 g 94 t

ORIGIN

Query Match 32.3%; Score 88.4; DB 33; Length 391;
Best Local Similarity 84.5%; Pred. No. 4.7e-11;
Matches 98; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 8 AGATTCTTCGCGCTCGCGTGAACACAAAGGGATCGGTCCTATATACCAATACCA 67
|||||
Db 102 AGATCCCGGGGACGGGTGAACACAAAGGGATCGGTCGCTATATACCAATACCA 161
|||||

QY 68 AACTTGATAATCTAGATTCTGTGTTCTGTATAGACCATGTTGTAGTAGG 123
|||||

Db 162 AACTGGACAAGAACTAGATTCTGTGTTCTACTTATATAGATCATGTTGTAGTAGG 217
|||||

RESULT 9

H09173/c 387 bp mRNA EST 23-JUN-1995
LOCUS y198h02.s1 Soares infant brain 1N1B Homo sapiens CDNA clone
IMAGE:46519 3', mRNA sequence.
H09173
ACCESSION H09173.1 GI:873995
VERSION H09173.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 387)

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 5, 1995 this sequence version replaced gi:797954.
Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1081
High quality sequence stops: 314
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1081 Std Error: 0.00
Seq primer: Promega -2ml13
High quality sequence stop: 314.

FEATURES

source
Location/Qualifiers
1..387
/organism="Homo sapiens"
/db_xref="GDB:419060"
/db_xref="taxon:9606"
/clone_lib="Soares infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH108 (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGAAATTCGCGCCGACGAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 132 a 67 c 72 g 112 t 4 others

ORIGIN

Query Match 29.9%; Score 81.8; DB 22; Length 387;
Best Local Similarity 85.7%; Pred. No. 1.5e-09;
Matches 150; Conservative 0; Mismatches 17; Indels 8; Gaps 5;

QY 108 CCATGTTTGTAGTAGTA-AGAGGAAACTTC--TATATTCTGAACAGCCCTACATT 163
|||||

Db 359 CCATGTTTGTAGTAGTANAGAGGAAACTTCCTATATCTCTGAACAGCCCTACATT 300
|||||

QY 164 TTACAAAATTTTGTAGTTCTTTTGTAGAGT--CTTATCTCTGTAG-CTATATACAGTTCA 220
|||||

Db 299 TTACAAAATTTTGTAGTTCTTTTGTAGAGTCTTTATCTGTAGCCTATATACAGTTCA 240
|||||

QY 221 TGT-CTGATTAGCATTTGTCACGAGTAAGCTGGAACATGAAAATTCAAAAT 274
|||||

Db 239 TGTCTGATTAGCATTTGTCACGAGTAAGCTGGAACATGAAAATTCAAAAT 185
|||||

RESULT 10

AA334229 274 bp mRNA EST 21-APR-1997
LOCUS EST38412 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA334229
ACCESSION AA334229
VERSION AA334229.1 GI:1986472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 274)

REFERENCE
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Val, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palmanques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrle, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl.), 3-174 (1995)

12140200

On Dec 3, 1996 this sequence version replaced gi:1119186.

Other ESTs: THC126123

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

Location/Qualifiers

1..274

/organism="Homo sapiens"

/db_xref="ATCC (inhost):135902"

/db_xref="taxon:9606"

/clone_lib="Embryo, 9 week"

/dev_stage="embryo, 9 wks"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT	87 a	51 c	53 g	81 t	2 others
------------	------	------	------	------	----------

ORIGIN

Query Match	27.2%	Score	74.6;	DB	32;	Length	274;
Best Local Similarity	98.7%	Pred. No.	7.4e-08;				
Matches	74;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						

QY	49	TGCCTATATCAACATACCAACTTGATAATATCTAGATCTGTGTCGTATTATAGAC	108
----	----	---	-----

Db	1	TGCCTATATCAACATACCAACTTGATAATATCTAGATCTGTGTCGTATTATAGAC	60
----	---	---	----

QY	109	CATGTTGTAGTAGG	123
----	-----	----------------	-----

Db	61	CATGTTGTAGTAGG	75
----	----	----------------	----

RESULT 11

AA488906/c

LOCUS

DEFINITION

aa55a02.r1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824810 5'

similar to TR:G607003 G607003 BETA TRANSDUCIN-LIKE PROTEIN. ;, mRNA

sequence.

ACCESSION

AA488906

VERSION

AA488906.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 EF from Amersham

High quality sequence stop: 446.

FEATURES

source

Location/Qualifiers

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:824810"

/clone_lib="NCI-CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73b-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI). Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I-oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	158 a	139 c	100 g	162 t
------------	-------	-------	-------	-------

ORIGIN

Query Match	23.6%	Score	64.6;	DB	34;	Length	559;
Best Local Similarity	98.5%	Pred. No.	1.3e-05;				
Matches	64;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						

QY	59	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	118
----	----	--	-----

Db	559	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	500
----	-----	--	-----

QY	119	GTAGG	123
----	-----	-------	-----

Db	499	GTAGG	495
----	-----	-------	-----

RESULT 12

T10350/c

LOCUS

DEFINITION

seq1091 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone

b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.

T10350

ACCESSION

T10350.1

VERSION

T10350.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 EF from Amersham

High quality sequence stop: 446.

FEATURES

source

Location/Qualifiers

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:824810"

/clone_lib="NCI-CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73b-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI). Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I-oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	158 a	139 c	100 g	162 t
------------	-------	-------	-------	-------

ORIGIN

Query Match	23.6%	Score	64.6;	DB	34;	Length	559;
Best Local Similarity	98.5%	Pred. No.	1.3e-05;				
Matches	64;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						

QY	59	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	118
----	----	--	-----

Db	559	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	500
----	-----	--	-----

QY	119	GTAGG	123
----	-----	-------	-----

Db	499	GTAGG	495
----	-----	-------	-----

RESULT 12

T10350/c

LOCUS

DEFINITION

seq1091 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone

b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.

T10350

ACCESSION

T10350.1

VERSION

T10350.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 EF from Amersham

High quality sequence stop: 446.

FEATURES

source

Location/Qualifiers

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:824810"

/clone_lib="NCI-CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73b-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided by Dr. Louis M. Staudt (NCI). Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I-oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	158 a	139 c	100 g	162 t
------------	-------	-------	-------	-------

ORIGIN

Query Match	23.6%	Score	64.6;	DB	34;	Length	559;
Best Local Similarity	98.5%	Pred. No.	1.3e-05;				
Matches	64;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						

QY	59	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	118
----	----	--	-----

Db	559	ACAAATACCAAACTTCAATATCTAGATCTGTGTCGTATTATAGACCATGTTTGA	500
----	-----	--	-----

QY	119	GTAGG	123
----	-----	-------	-----

Db	499	GTAGG	495
----	-----	-------	-----

RESULT 12

T10350/c

LOCUS

DEFINITION

seq1091 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone

b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.

T10350

ACCESSION

T10350.1

VERSION

T10350.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 EF from Amersham

High quality sequence stop: 446.

FEATURES

source

Location/Qualifiers

1..559

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:824810"

/clone_lib="NCI-CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73b-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-),

provided

```

Tel: 2129602313
Fax: 2127813577
Email: cue@cuccfa.ccc.columbia.edu
antisense, Ampicillin
Seq primer: M13 Forward (Universal).
Location/Qualifiers
1..259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="b4HB3MA-CO8-HAP-Ft287"
/clone_lib="b4HB3MA COt8-HAP-Ft"
/lab_host="E. coli"
/note="vector: Laimid BA; Site_1: Not I; Site_2: Hind III;
Size-selected cDNA from polyA+ RNA from human brain.
3-month old neonate. This library is the result of an
attempted normalization of library b4HB3MA."
90 a 50 c 46 g 72 t 1 others
BASE COUNT
ORIGIN

Query Match 23.4%; Score 64.2; DB 20; Length 259;
Best Local Similarity 93.9%; Pred. No. 1.8e-05; Indels 1; Gaps 1;
Matches 77; Conservative 0; Mismatches 4;

QY 194 CTTATCCTGCTAGCTATATACAGTCATGCTGTGATTTAGCATTTGTT-CACGAGTAAAGC 252
+
Db 259 CTTATCCTGCTAGCTATATAGCAGTTCATGCTGTGATTTAGCATTTGTTTACACGAGTAAAGC 200
+
QY 253 TGGAACTATGAAATTTGAAAT 274
+
Db 199 TGGAACTATGAAATTTGAAAT 178
+

RESULT 13
LOCUS AQ587717
DEFINITION AQ587717.1 TF CITBI-E1 Homo sapiens genomic clone 2646F4,
genomic survey sequence.
ACCESSION AQ587717
VERSION AQ587717.1 GI:5014397
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
I (bases 1 to 596)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
Unpublished (1997)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2646F4"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
FEATURES
source

```

Search completed: April 5, 2000, 11:35:51
Job time: 22828 sec

